

QY	550	agggacacagcttcagagatctctgatacttcttcgggagaccattttaactgataat	609
Db	1	aggacacacagcttcagagatctctgatacttcttcgggagaccattttaactgataat	60
QY	610	tgcacaatgttcagatcttcctctctgtaatgcagaatgtgcacaagtcttcggaatttcg	669
Db	61	tcacacaaatgttcagatcttcctctctgtaatgcagaatgtgcacaagtcttcggaatttcg	120
QY	670	tctgttcacagcatcagacagaaatggtgcagaatcctttggaacaactgcgtcagatggg	729
Db	121	tctgttcacagcatcagacagaaatggtgcagaatcctttggaacaactgcgtcagatggg	180
QY	730	gcagaaacttciaaattttctttaaatacagagaaaaaattcattttgtaaatagataag	789
Db	181	gcagaaacttciaaattttctttaaatacagagaaaaaattcattttgtaaatagataag	240
QY	790	agtaactggagctcctaattgtctgtatcagtaaaagggcaaaatacacagatggagacaga	849
Db	241	agtaactggagctcctaattgtctgtatcagtaaaagggcaaaatacacagatggagacaga	300
QY	850	ttgaaaagggaan--caattcatgcactctgtataatgataagtttctctaaagaagggcagc	907
Db	301	ttgaaaagggaan--caattcatgcactctgtataatgataagtttctctaaagaagggcagc	360
QY	908	tcaaagacataccag	922
Db	361	tccaagacataccag	976
RESULT 5	AA897178/c		
LOCUS	am9e08.s1 Soares_NFL.T.GRC_S1 Homo sapiens	238 bp	linear
DEFINITION	IMAGE:1466342.3', mRNA sequence.		EST 04-JAN-1998
ACCESSION	AA897178		
VERSION	AA897178.1	GI:3033798	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
COMMENT	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP). Tumor Gene Index Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cga@bbs.femail.nih.gov This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert length: 847 Std Error: 0.00 Seg primer: -40m3 fwd. ER from Amersham High quality sequence stop: 132. Location/Qualifiers		
FEATURES			
source			

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1466342"
 /clone_1lib="Scarses_WFL_T_GBC_S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pRT73D-Pac (Pharmacia) with
 a modified polylinker; Site1: Not I; Site 2: Eco RI;
 Equal amounts of plasmid DNA from three normalized
 libraries (fetal lung NBH19F, testis NHT, and B-cell
 NCL-GAP-GCB1) were mixed, and se circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from pools of 5,000 clones made
 from the same 3 libraries. The pools consisted of
 I.M.A.G.E. clones 297480-302087, 682632-687229,
 726608-728711, and 729096-731397. Subtraction by Bento
 Soares and M. Fatima Donaldso. "

BASE COUNT	71 a	53 c	45 g	69 t
ORIGIN				
Query Match	16.1%;	Score 226;	DB 9;	Length 238;
Best Local Similarity	99.6%;	Pred. No. 4.5e-47;		
Matches 237;	Conservative 0;	Mismatches 0;	Indels 1;	Gaps 1;
QY 1156	actggagctgcattcattatgtaattgagccattgtgacatgtaattcagccagaagaacctg	1215		
DB 238	actgggcctccatttcattatgtgaatggccaccttggcactgatgattcagccaggaagaaacctg	179		
QY 1216	tggcgctgatagataaaggagttggtgttcgaagtgcactggaataacaagtttaatgcacga	1275		
DB 178	tggcgggaatgaatgaatgaatggatggtgttcacaaactgcgaatgaacagtttaatgacacga	119		
QY 1276	taataatgcctccgtcatalg- ⁴ tgaattgctcaataaaggttaaaaaacggtattccctggt	1334		
DB 118	tttaattagctcctgcataatg ⁴ taaatgtcctaataacagtttaaaaaacggtatttccctggt	59		
QY 1335	tctgatccacaactatcaagtgtctacagaacacctatcatcgttctgtaagaacct	1392		
DB 58	tctgatccacaactatttcaactgtttacagaacacctcattatcagttgtaagaacacct	1		

RESULT 6
AA776169
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AA776169 209 bp mRNA linear EST 05-FEB-1996
ae80c02.s1 StrataGene schizo brain SII Homo sapiens CDNA clone
IMAGE:790466 3', mRNA sequence.
AA776169
AA776169.1 GI:2835503
EST.
human.
Homo sapiens
Euryarchaea; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 209)
Hiller, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Kizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin,
J., Moore, B., Schellenberg, K., Stepec, M., Tan, F., Theisling, B.,
White, Y., Wylie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone; polyr not found
Seq primer: -40m13 fwd. RT from Amersham
High quality sequence stop. 206.
Location/Qualifiers
1..209

1. 209
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone IMAGE:970466"
/clone_id="Stratagene schizo brain S11"
/sex="male"
/tissue_type="schizophrenic brain S-11 frontal lobe"
/dev_stage="34 years old"
/lab_host="SOLR (Kanamycin resistant)"
/note="Vector: Bluescript SK-; Site_1: EcoRI; Library
constructed from S-11 frontal lobe, male, 34 years old,
50% caucasian, 50% Aleutian. Schizophrenic suicide.
Random primed into EcoRI site of ZAP II Vector. Mass
excised. Custom library. Avg insert length 1.4kb.
Material obtained by Johnston N., Torrey, E.F., Yolken R.,
and the Stanley Neuropathology Consortium - Analysis of
RNAs from the Brains of Individuals with Psychiatric

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 05:31:11 ; Search time 3619.39 Seconds
(without alignments)
5224.426 Million cell updates/sec

Title: US-09-693-205-7_COPY_5300_6700

Perfect score: 1401
Sequence: 1 gtagcagtaaacactagagc.....aaagacacttaagaagt 1401

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Archived: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estbda:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estcpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pin:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	752.2	53.7	1083	10	BM476887 AGENCOURT
2	673.4	48.1	795	10	R17106 EST20108 C1
3	458.4	32.7	1010	10	BM472954 AGENCOURT
4	345.2	24.6	376	9	AL596565 DFF2P451L
5	226	16.1	238	9	AA897178 am09608.s
6	203.4	14.5	209	9	AA776169 aa80c02.s
7	142.4	10.2	148	9	AA776670 aa80c02.s
8	134.6	9.6	499	12	A2177954 SP_0148_B
9	114.8	8.2	505	9	AM295072 UI-H-BW0
10	107.8	7.7	434	9	AM470182 xw61a11.x
11	87.6	6.3	1067	12	CNS02W80 Tetradon
12	87.2	6.2	586	10	BF650306 NF087D1E
13	85	6.1	918	12	A2136287 SP_0169_B
14	84	6.0	807	12	A2185191 SP_1004_B
15	80	5.7	431	12	A2182120 SP_0188_A
16	79.6	5.7	331	9	A1685137 wa75a03.x
17	67.2	4.8	821	12	BH601255 BOGYW28TR

18	61.6	4.4	643	9	BB485245
19	58.4	4.2	284	9	BB254184
20	58.2	4.2	445	10	BF646344
21	57.8	4.1	616	12	A2963183 2M0232F17
22	55.2	3.9	234	9	BB254737
23	51.4	3.7	721	12	AQ254586
24	48.4	3.5	1101	12	CNS0039G
25	47.8	3.4	1069	12	CNS055ESA
26	47.2	3.4	936	12	CNS02076Y
27	46	3.3	959	12	CNS040NC
28	44.8	3.2	1101	12	CNS0006J
29	44.4	3.2	555	12	AQ310011
30	44.4	3.2	560	12	AQ310013
31	43.8	3.1	937	12	CNS040H
32	42.6	3.0	380	9	AM517860
33	42.6	3.0	407	9	AM294501
34	42.6	3.0	548	10	BM274569
35	42.6	3.0	609	10	BE440138
36	42.4	3.0	987	12	CNS014PO
37	42.2	3.0	987	12	CNS014PO
38	42	3.0	979	12	CNS00D9H
39	41.8	3.0	536	12	AQ757450
40	41.8	3.0	782	9	AL552607
41	41.6	3.0	496	12	BM479411
42	41.6	3.0	524	12	BH600863
43	41.6	3.0	547	12	BH502803
44	41.6	3.0	611	12	BH436907
45	41.6	3.0	616	12	BH463467

ALIGNMENTS

RESULT 1
BM476887 1083 bp mRNA linear EST 05-FEB-2002
AGENCOURT 6481789 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5555441
LOCUS 5', mRNA sequence.
ACCESSION BM476887.1 GI:18525929
VERSION BM476887
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1083)
NIH-MGC http://mgs.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12275 row: j column: 18
High quality sequence stop: 696.
Location/Qualifiers
1. 1083
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5555441"
/clone_lib="NIH_MGC_71"
/tissue_type="leiomysarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb.
BASE COUNT 334 a 219 c 212 g 318 t
ORIGIN

THIS PAGE BLANK (USPTO)

TITLE	Study of expressed sequences tags in adipose tissue 1995
JOURNAL	unpublished (1995)
COMMENT	contact: Frederic Boullaud Centre de Recherche sur l'Endocrinologie moléculaire et le Développement

FEATURES	Location/Qualifiers
source	1. .795

BASE COUNT	243 a	150 c	166 g	216 t	20 others
ORIGIN					

Local Similarity 34.4%; Pred. NO. 3.98-161;
Matches 743; Conservative 1; Mismatches 30; Indels 13; Gaps 5

Db 9 ATCCAGATCTNGGAAAAAGCCACGAAAGAGGAAATCCTTTAAACIGGACAGTATGGA 68

DB 69 ATTAGGATTCATTCTGTGTATCATATTCACAGACTGCCCATCTTTATTTCTGGCATGAC 128

[illegible]

OY 577 attatctggaaccattttaactqatatlqacagatattcaatttccttccat 636
.....TTCGCTCAGATGTTTGCGA 248

637 aatgcagaatgycaaagtcttgcgaatttcgctctgttccagcatcagacacaatgtgc 696

697 cagaatctttgg--acaactgcgctcagat--gggacagaactctatgtttcttaa 752

753 tcac-atgcaaaaatttctattgtg----aaatgataagtagtactgagctctaa 806

863 aataaggaattctgcgcgcgtatc---cagtaaaagcgcaaaatccacagatcggagacggaatc 860

[illegible]

609 TCACAAATACCTATACCTATGGATACGAGGACTCTGNAGCAATCTTACTACGTGGCT 668

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 06:53:41 ; Search time 3530.57 Seconds

(without alignments)
3514.853 Million cell updates/sec

Title: US-09-693-205-7_COPY_12200_12792

Perfect score: 593

Sequence: 1 acattctatgtttacaggct.....accctcacattttatgctt 593

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Archived: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database :

GenEmbl:
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_ov:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_inv:*
32: em_hlg_other:*
33: em_hcgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	593	100.0	4318	9	AB018273	AB018273 Homo sapi
2	593	100.0	12793	6	AX119931	AX119931 Sequence
3	593	100.0	12793	9	AF193556	AF193556 Homo sapi
4	593	100.0	92693	9	AL157766	AL157766 Human DNA
5	593	100.0	99819	2	AC079761	AC079761 Homo sapi
6	426.6	71.9	3289	9	AB056815	AB056815 Macaca fa
7	327.4	55.2	418	11	G36555	G36555 SHGC-53325
8	61	10.3	169794	2	AC004688	AC004688 plasmodu
9	61	10.3	196149	2	AC004709	AC004709 plasmodu
10	59.6	10.1	4601	3	DM01584	DM01584 Drosophila
11	59.6	10.1	19517	3	DM037541	DM037541 Drosophila
12	57.4	9.7	12029	3	AE001426	AE001426 plasmodu
13	57.2	9.6	60604	2	AC023466	AC023466 Homo sapi
14	57	9.6	183813	2	AC012204	AC012204 Homo sapi
15	57	9.6	197225	2	AC093835	AC093835 Homo sapi
16	56.8	9.6	110000	2	PFMAL4P1_2	PFMAL4P1_2 Continuation (3 of
17	56.6	9.5	11691	6	AX347143	AX347143 Sequence
18	56.6	9.5	349980	6	AX344564	AX344564 Sequence
19	55.4	9.3	124635	6	AP000593	AP000593 Homo sapi
20	55.2	9.3	1815	1	AF267220	AF267220 Candidatu
21	35.2	9.3	162075	9	HS127D3	HS127D3 Homo sapi
22	55	9.3	1141	6	AX083744	AX083744 Sequence
23	55	9.3	13038	6	AX346176	AX346176 Sequence
24	54.8	9.2	145598	6	AC008132	AC008132 Homo sapi
25	54.8	9.2	15387	6	AX345086	AX345086 Sequence
26	54.6	9.2	4611	3	PEA132006	PEA132006 plasmodu
27	54.4	9.2	7918	8	AF325123	AF325123 Arabidops
28	54.4	9.2	13574	6	AX346218	AX346218 Sequence
29	54.4	9.2	62268	9	HS1178121	HS1178121 Human DNA
30	54.4	9.2	141748	8	AC009333	AC009333 Arabidops
31	54.4	9.2	187048	2	AC024921	AC024921 Homo sapi
32	54	9.1	349980	6	AX344558	AX344558 Sequence
33	54	9.1	349980	6	AX344559	AX344559 Sequence
34	53.8	9.1	863	11	CNS06EVO	CNS06EVO T7 end of
35	53.8	9.1	189790	2	AC107420	AC107420 Homo sapi
36	53.4	8.9	186625	9	AC006487	AC006487 Homo sapi
37	53	8.9	141469	9	AL139811	AL139811 Human DNA
38	53	8.9	161384	2	AC083801	AC083801 Homo sapi
39	52.8	8.9	17848	6	AX277865	AX277865 Sequence
40	52.8	8.9	17848	6	AX323550	AX323550 Sequence
41	52.8	8.9	17848	6	AX348363	AX348363 Sequence
42	52.8	8.9	19087	6	AX345694	AX345694 Sequence
43	52.8	8.9	77835	2	PFMAL13P2_3	PFMAL13P2_3 Continuation (4 of
44	52.6	8.9	150754	9	AC023481	AC023481 Homo sapi
45	52.6	8.9	175358	9	AC007981	AC007981 Homo sapi

ALIGNMENTS

RESULT 1
AB018273 4318 bp mRNA 11linear PRI 16-JUN-1999
LOCUS Homo sapiens mRNA for KIAA0730 protein, partial cds.
AB018273
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE

AB018273.1 GI:3862180
Homo sapiens adult male brain cDNA to mRNA, clone_11b:pb1uescript11
SK plus clone:hk03632.
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Nagase, T., Ishikawa, K., Suyama, M., Kikuno, R., Miyajima, N.,
Tanaka, A., Kotani, H., Nomura, N. and Ohara, O.
Prediction of the coding sequences of unidentified human genes. XI.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro
DNA Res. 5 (5), 277-286 (1998)
99087487
2 (bases 1 to 4318)
Ohara, O., Suyama, M., Nagase, T., Ishikawa, K. and Kikuno, R.
Direct Submission

THIS PAGE BLANK (USPTO)

JOURNAL
Submitted (08-0CR-1998) Osamu Ohara, Kazusa DNA Research Institute
Laboratory of DNA Technology, Yana 1532-3, Kistazazu, Chiba
292-0812, Japan (E-mail: cdnainfo@kazusa.or.jp, Tel.: +81-438-52-3913
Fax: +81-438-52-3914)

FEATURES	Location/Qualifiers
source	1. .4318

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hx03632"
/sex="male"
/rassue_type="brain"
/clone_lib="pbluescriptit sk plus
/dev stage="adult"
1..3015
/gene="KIRA0730"
<1..3015

```

/gene="K1AA0730"
 /codon_start=1
 /product="K1AA0730 protein"
 /protein_id="BA034450.1"
 /db_xref="GI:3882181"
 /translation="VFVMTSVWEFLRNIGLAKYILSSQOOLQPAKEISVRANTEENKSI
 TQNTQDILLHHIFQEBMDLGSNFLKELSLIPPCPEAPAEPIRFHQYQVENVGTG
 PLIKNGAQNPKFKOCQDYLQILMLTSCPTLPKCAPISTKBQSGDQGBQEDQVIL
 MNLNIDPDLKVIINCRNCITNTITLDEMPTRAKVLKSLTSEIETSAEKRETFQDVR
 VAFVWEDGDKLLKPEEVVINYLESDFEPVLYIKLPLELGTGHQFLKHLGTDIISTQ
 QVEVLSIRIFKNSGKQLPDNEMRYKRVSGLSFSDNLQSVKSNDELVNADLYL
 PQDGLGVKSLSEVFDAPHYKVRSGIQTGQVOMLVDSQCYGKHQHPGKTLIMLP
 KAPRLSLISILEQDLEMPKVCQCALCSLOGRQLQILSSQPTIGLIRIKKEND
 AFLANBEKAIKCKLAREGLKSCERKQLTTLRVGENTIPRKSETFAPLRFENGA
 ILLYIQHSQSKDINFLALAMTLKSNATDMLSDTSLYLAIMGCDNIYRAGELNLSG
 KDDSPSKLELPMDGTPIPAEIHYYTLMDPNVNNYPCGYGLVLADESGDYSYQI
 TTYVAIIYQVEEREDADNNSFLGIYQIDIGYSEKVIYSLVLKFRSPREESQDSD
 ASPTSPPEFLIPGLRSIPLPLSGRSHSTKSHQSPKLLINSLPELKEVTSYVVI
 QAMKLPESRRKTIIRRLYIKMHPDKNPEHNDIANEYFKHLQNEINBLEKQATLDNNA
 RSKRSTFSTSASRFQSDKYSFORFTYSNMQNTSHKSEKQOONKEKCPSPSAQYTS
 FVPPFTKSVGNVBARWLROARNFSAARNDLKHNAEMWCFCFYSTKALIAALAA
 VYARKSDKDVPLAOKIIEEYSOOLETITVHTLEAVGYDVSILKTRKTDLPDLPQ
 PNDRTSEVAMVMECTACIILIKLENMQOK"

Query Match	100.0%	Score 593	DB 9	Length 4318
Best Local Similarity	100.0%	Pred. No. 4,7e-92		
Matches 553	Conservative 0	Mismatches 0	Indels 0	Gaps 0
1	acactctatgtttacagagcttcctgctttgtatgtagatagcaacggaanaactcaaatgtg	60		
b	ACATCTTAATGTTTACAGAGCTTCTGTTTGATGAGATAGCAACGGAAACTCAAAATGGT	3709		
61	ggcagctctatattacacagttgttagtattgtttcttctgaaactgcttgcagaagaacaatt	120		
b	GCGAGTCTTATATACAGAGTGTAGATATGTTCTCTGGAACCTGCTGCCAAGACAACTAT	3769		
121	tattaaactgttagaacaacttgcttatgtttgtgtgtatcaatattccacaagtgtata	180		
b	TATTAAGTGTAAACAACCTGCTTATGTTGTGTGACATATTTTCCAAATGTTATA	3829		
181	attatatagtgtgtgttgaaacagatgcaactctttgttgcctaaagtgctcagttaa	240		
b	ATTTAATATAGTGTGTGAACAGGATGCAATCTTTGTTGTTCAAGAGCTGCTCAGATTAA	3889		
241	aaaaaaaacaaactttcttcaatatggaatgtagttagtatttttaactttaaaac	300		
b	AAAAAAAACAACTTTTCTTCAATATGGCATGATGAGGGATTTTAACTTTAAAAAC	3949		
301	atcaaaaattgtttaaaatcaatgtgtatcttagtatttaaatatcggctataatc	360		
b	ATCAAAAATTGTTAAATCACTTTGTGTTATCTAGTAGTTTAATTAATCGGCTTATATTC	4009		
361	cccatgaatgatcagaactgacattaatcaatgattgttcctgcacatgcttcttaactt	420		

Db	4010	CCCATGAATGATGAGAACGACGACATTAAATCAATGTTTGTCGGCATCCTCTTACTT	4069
QY	421	aacataatcctcttgcagaagtaaaaggtlaagtaaatgattatataaaglyactg	480
Db	4070	AACATATTTCTTTTGCAGATGTAAAGATATGATTAATTAATTAATAGTATCGG	4129
QY	481	cgttaaatgatgcttaataatacttataatgaatlaaagggttaagaaatgttgaact	540
Db	4130	CTGTAAATGATGCTTAATTAATTAATTTATGCAATTAAGGGCTTACAGACATGTGAACTT	4189
QY	541	tttttaactttatcttggaataaggaatgtttgcactccaatttatatgtc	593
Db	4190	TTTTTACTTTTATTTGGGAATAGGAATGTTTCACCTCCACATTTTATVGGCT	4242

RESULT	2			
AX119931	LOCUS	12793 bp	DNA	linear
AX119931	AX119931	Sequence 1	from Patent WO0129266.	
AX119931	AX119931	ACCESSION		
AX119931.1	GI:14036678	VERSION		

KEYWORDS	SOURCE
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE	1 (bases 1 to 12793)
JOURNAL	Hudson T.J., Engert J. and Richter A.
FEATURES	Identification of asac mutations and methods of use therefor
	Patent: WO 0129266-A 1 26-AR-2001.
	McGILL UNIVERSITY (CA); Hopital Sainte-Justine (CA)
	Location/Qualifiers

```
BASE COUNT      4163 a      2256 c      2487 g      3887 t
ORIGIN          /organism="Homo sapiens"
                /db_xref="taxon:9606"
```

Query Match	100.0%	Score 593	DB 6	Length 12793
Best Local Similarity	100.0%	Pred. No. 3.6e-92		
Matches 593	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy 1	acactctatgcttcacagctccctcgtcttcgtacgaagatagcaacggaacaaactcaaatggt	60		
Db 12201	ACACCTTATGTTTACAGGCTCCGTGTTGGANGAAGATAGCAACGGAAAACTCAAAATGTT	12260		
Qy 61	ggcagctctctattaccagctgttagtattgttctcgaaactgcttccagaacaacatt	120		
Db 12261	GGCAGTCTTATTTACCAAGTTGTATGTTATGTTTCGGAAACGTCTTGGCAGAACACACTT	12320		
Qy 121	tattaactgttaagaaccttgcttattgtttgtgtatcatatttccacaaatttata	180		
Db 13321	TATTAACTGTTAAGAACACTTCTTATGTTGTGTGTCATATTTTCCACAAATGTTATA	12380		
Qy 181	attatatagtgtgttgaacagatagcaactcttgtcttcctaaagttgcgcagttaa	240		
Db 12381	ATTATATAGTGTGTGTGAACAGATGCAATCTTTGTGTCTTAAAGGTGTCGCAAGTTAA	12440		
Qy 241	aaaaaaaaaacaccccttcttccaatatgcatgttagtgaagtttcttcaactttaaaac	300		
Db 12441	AAAAAAAAACAACCTTTTCTTTCATATATGAGCAGTAGTGGAGTTTAAAACTTTAAAAAC	12500		
Qy 301	atcaaaaatgtttaaatcatcttgcttaccagtagtttaataatatacggcctatacttc	360		
Db 12501	ATCAAAAATGTGTTAAATTCATGTTGTATATCTAGATGTTTAAATTAATGCGGCTTAATTTTC	12560		
Qy 361	cccatgaatgtagcaaacatgacatttaactatcatgttttccgcgcagtcctcttaactt	420		
Db 12561	CCCATGAATGATCGAAGCTGACATTTAATTATGTTTGTCTCGCCATGCTTCTTACTTT	12620		
Qy 421	aaacattctcttcgcagaaagtataaagtgtaataatgaattagttatataaagtgtaag	480		

THIS PAGE BLANK (USPTO)

Wed May 22 09:23:25 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model
May 22, 2002, 05:31:11 : Search time 3619.39 Seconds
(without alignments)
5224.426 Million cell updates/sec

Run on:
US-09-693-205-7_COPY_5300_6700
Title:
1 gtagcagtaactaagagc.....aaagcacttaagaagt 1401
Sequence:
IDENTITY NUC 670
Gapop 10.0, Gapext 1.0
Scoring table:
13736207 segs, 6748477542 residues
27472414

Searched:
Total number of hits satisfying chosen parameters:
Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gD_est12:*
10: gD_est1:*
11: gD_hic:*
12: gD_gss:*
13: em_gss_hum:*
14: em_gss_hiv:*
15: em_gss_pln:*
16: em_gss_vitl:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Match	Length	DB	ID	Description
1	752.2	53.7	1083	10	BM476887	BM476887 AGENCOURT
2	673.4	48.1	795	10	R17106	R17106 EST20108 C1
3	458.4	32.7	1010	10	BM472954	BM472954 AGENCOURT
4	345.2	24.6	376	9	AL596565	AL596565 DKFZP451L
5	203.4	16.1	238	9	AA897178	AA897178 am09608.s
6	144.4	10.2	148	9	AA776670	AA776670 ae08002.s
7	134.6	9.6	499	12	A2177954	A2177954 SP_0148.B
8	114.8	8.2	505	9	AA295072	AA295072 UI-H-BM0
9	107.8	7.7	434	9	AA470182	AA470182 xw61a11.x
10	87.6	6.3	1067	12	CNS02W80	AL216729 Tetradon
11	87.2	6.2	586	10	BF650306	BF650306 NF087D11E
12	85	6.1	918	12	A2136287	A2136287 SP_0169.B
13	84	6.0	807	12	A2185191	A2185191 SP_1004.B
14	80	5.7	431	9	A2182120	A2182120 SP_0188.A
15	79.6	5.7	331	12	AT685137	AT685137 wa75a03.x
16	67.2	4.6	821	12	BH601255	BH601255 BCGY28TR

us-09-693-205-7_copy_5300_6700.std.rst

18	61.6	4.4	643	9	BB485245	BB485245
19	58.4	4.2	284	9	BB254184	BB254184
20	58.2	4.2	445	10	BF646344	BF646344
21	57.8	4.1	616	12	A2963183	A2963183
22	55.2	3.9	234	12	BB254737	BB254737
23	51.4	3.7	721	12	AA254586	AA254586
24	48.4	3.5	1069	12	CNS00396	AL130281 Tetradon
25	47.8	3.4	936	12	CNS02T6Y	AL1212803 Tetradon
26	47.2	3.3	959	12	CNS040NC	AL062049 Drosophila
27	46	3.2	1101	12	CNS0006J	AL0310011 CTRB1-E1
28	44.8	3.2	555	12	AA310013	AA310013 Tetradon
29	44.4	3.2	560	12	AA310013	AA310013 Tetradon
30	44.4	3.1	937	12	CNS04HOH	AA294501 UI-H-B12
31	43.6	3.0	480	9	AA517860	AA517860 PEST00A4
32	42.6	3.0	508	10	BM274569	BM274569 PEST00A4
33	42.6	3.0	609	12	BE440138	BE440138 HTM1-946A
34	42.6	3.0	670	12	AA308275	AA308275 CTRB1-E1
35	42.4	3.0	987	12	CNS014PQ	AL104456 Drosophila
36	42.2	3.0	979	12	CNS00DPH	AL060395 Drosophila
37	42.2	3.0	536	12	AA757450	AA757450 HS-5466-B
38	42	3.0	782	9	AL535607	AL535607 AL535607
39	41.8	3.0	496	12	BH479411	BH479411 BCGXK20TF
40	41.6	3.0	524	12	BH600863	BH600863 BCGXK30TF
41	41.6	3.0	547	12	BH502803	BH502803 BCGXK30TF
42	41.6	3.0	611	12	BH436907	BH436907 BCGXK30TF
43	41.6	3.0	616	12	BH463467	BH463467 BCGXK40TF
44	41.6	3.0				
45	41.6	3.0				

ALIGNMENTS

RESULT 1
BM476887 1083 bp mRNA linear EST 05-FEB-2002
AGENCOURT_6481789 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5555441
5', mRNA sequence.
BM476887
BM476887.1 GI:18525929
EST.
Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (base http://mgi.nci.nih.gov/)
NIH-MGC National Institutes of Health, Mammalian Gene Collection (MGC)
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@nci.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLAM12275 row: 1 column: 18
High quality sequence stop: 656.
Location/Qualifiers
1. 1083
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5555441"
/clone="NIH_MGC_71"
/clone="lib=telomovosarcoma"
/tissue_type="DH10B (phage-resistant)"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-Sport6; Site: 1; NOTI
Site: 2; Salt: cloned unidirectionally. Primer: Oligo dT
Average insert size 2.1 kb.
334 a 219 c 212 g 318 t

FEATURES
source
BASE COUNT
ORIGIN

us-09-693-205-7_copy_5300_6700.std.rst

```

CNA Library Arrayed by: The Bioscience Corporation " can be
DNA Sequencing by: Agencourt Clone Distribution Information at:
Clone distribution: MGC clone distribution information at:
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.lnl.gov
LML12323 row: h column: 13
High quality sequence stop: 738.
Location/Qualifiers
1..1010
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5574588"
/clone_1ib="NIH_MGC-88"
/issue-type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant), Vector: pCMV-Sport6;
/note="Organ: small intestine; Location: undirectionally;
site_1: NotI. Site_2: SalI; Cloned unidirectionally;
cldg-dr primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."
BASE COUNT      293 a          209 c          191 g          317 t
ORIGIN
Query Match      32.7%; Score 458.4; DB 10; Length 1010;
Best Local Similarity 99.4%; Pred. No. 3e-106; 1; Indels 2; Gaps
Matches 481; Conservative 0; Mismatches
QY    920   cagtttaacaataactatctactatgatgtacgtggagccttgaagaatacttactactgt 979
       |||||
Db    20    CAGTTCAACAATAATTAACCTTACTATGATGCTACTGTGAGCGACTGTGAAGAANAATCTTACTACGT 79
       |||||
QY    980   ggctaatttgtaatagatcgaggctttcaagtatggagaaagatatctaaagtgataat 1039
       |||||
Db    80    GGCTAATTGGTAATNGATTCAGCGCTTTTCAGTAGTGAGAAAGTATCTTAAAGTGTCAATF 139
       |||||
QY    1040  cagctcacagaagaccagatatactctcttccacagtggtggagtagctgcctgcaata 1099
       |||||
Db    140  CAGCTCACAGAAGACCAGATATTTACTCTTTCCAGCTGAGTGTGAGTAGCTGCTGCATTA 199

```

[illegible]

```

FEATURES
source
    This is the 5' sequence of the Human Genome Project. No 31 sequence
    clone from S. Wiemann, Molecular Genome-Electronics, the CDNA
    Research Center (DKFZ), Email s.wiemann@dkfz.de, within the
    sequenced consortium of the Human Genome Project.
    This clone (DKFZ451L0110) is available at the RZPD in Berlin.
    Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
    Berlin-Charlottenburg, Germany; Email: clone@rzpd.de.
    Location/Qualifiers
        1..376
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="DKFZ451L0110"
        /clone_lib="451 (synonym: hlcl1) spinal cord"
        /tissue_type="human spinal cord"
        /tissue_stage="adult"
        /dev_stage="DH10B"
        /lab_host="psport1; Site_1: NotI; Site_2: SalI"
        /note="vector: 79 g 107 t 63 c
BASE COUNT
    127 a
ORIGIN
    24.6%; Score 345.2; DB 9; Length 376;
    98.4%; Pred. No. 1.9e-77; Indels 3; Gaps
    0; Conservative Matches 370;

```

`/db_xref="taxon:9606"`
`/clone_id="IMAGE:1466342"`
`/clone_lib="Soares_NFL_T_GBC_S1"`
`/lab_host="DH10B"`
`/note="Organ: pooled: Vector: pF73D-Pac (Pharmacia) with:
 Equal amounts of plasmid DNA from three normalized
 libraries (fecal lung NBH19W, testis NHT, and B-cell
 NB1-CGAP-GCB1) were mixed, and ss circles were made
in vitro. Following HAP purification, this DNA was used
 as PCR-amplified cDNAs from pools of 5,000 clones made
 from the same 3 libraries. The pools consisted of
 1.M.A.G.E. clones 297480-302087, 662632-687239,
 726408-728711, and 729096-731399. Subtraction by Bento
 Soares and M. Fatima Bonaldo."`

```

/db.xref="taxon:9606"
/clone="IMAGE:970466"
/clone.lib="Stratagene schlzo brain s1l"
/sex="male"
/tissue.type="schizophrenic brain S-11 frontal lobe"
/dev.stage="34 years old"
/lib.host="SOLR (knamycin resistant)"
/notes="Vector: Bluescript SK-; Site_1: EcoRI. Library constructed from S-11 frontal lobe, male, 34 years old, 50% caucasian, 50% Alutian. Schizophrenic suicide, excised. Custom library. Avg insert length 1.4kb. Material obtained by Johnson N., Torrey, E.F., Yolken R., and the Stanley Neuropathology Consortium - Analysis of RNAs from the Brains of Individuals with Psychiatric

```

FEATURES	source	Location/Qualifiers
	1. .499	/Organism="Strongylocentrotus purpuratus" /db_xref="taxon:7668" /clone="plate=148 Col=12 Row=N" /clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library" /note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli DH10B"
BASE COUNT	140 a	96 c 116 g 146 t 1 others
ORIGIN		
Query Match	9.6%	Score 134.6; DB 12; Length 499;

TITLE		National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	
JOURNAL		Unpublished (1997)	
COMMENT		Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Life Technologies catalog #: 11548-013 DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNLW at: www-bio.lnlw.gov/bdip/image/image.html Seq primer: -40up from Gidco High quality sequence stop: 412.	
FEATURES		Location/Qualifiers	
source		1..434	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/clone="IMAGE:2832476"	
		/clone_lib="NCI CGAP, Pan1"	
		/tissue_type="adenocarcinoma"	
		/lab_host="DH10B"	
		/note="Organ: pancreas; Vector: pCMV-SPORT6, site_1: SalI; site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.72 kb. Life Technologies catalog #: 11548-013"	
BASE COUNT		141 a 76 c 101 g 116 t	
ORIGIN			
Query Match 7.7%; Score 107.8; DB 9; Length 434; Best Local Similarity 58.6%; Pred. No. 7.5e-17; Matches 187; Conservative 0; Mismatches 132; Indels 0; Gaps 0;			
QY	122	ttaaagagcatcccttaatgcatatcccttcgtgaaagaaagtgtgaaagagcttcttcaaa	181
DB	115	TCAAGACATTTTGGAGAGATATCCAGAAGAGGAGCAGACTTCTTAAGGAATTAATTCA	174
QY	182	atgctgtagtgcgaagcgacagaacatctgtttgttgttgaaccttgacagcatccag	241
DB	175	ATGCGAAGATCTCTGGCGACGAGAAAGTTAAATTTTATATGATGAACATACGAGAA	234
QY	242	ttgatagatatattgatgataagtgaggcccatctgcaaggcgccagcacttgtgtgtaca	301
DB	235	CAGACACTTTTGGTCAAAAGATATGGCGCCATATCAGGGGCGCAGCTCTTATGTGTACA	294
QY	302	acaacagccattacagagaatgtagtgcgaagaaattcaagaatcttgaaaaagcgaca	361
DB	295	ACAAAGCGGTTTTCACCCGAGGAGACTGGCAGCGCATTCACAAATATAGCAACAACGAGA	354
QY	362	aagaggaatccctataaactgacagcatggaatggattcaatctgtgtcatca	421
DB	355	AAAAGATGATCTCTGGAAGTGGGAATTTGGATTGGTTAATTCGTCTATCATTA	414
QY	422	tcaagactgcgccacttt	440
DB	415	TAAAGATGTCTCTGTAT	433
RESULT 11 CNS02W80 1067 bp DNA linear GSS 15-MAR-2000 LOCUS CNS02W80/c DEFINITION Tetradon nigriviridis genome survey sequence PUC-Or1 end of clone 1/7609 of library G from Tetradon nigriviridis, genomic survey sequence. ACCESSION AL216729.1 GI:7875548 VERSION AL216729 KEYWORDS GSS: genome survey sequence. SOURCE Tetradon nigriviridis. ORGANISM Tetradon nigriviridis Tetradon nigriviridis Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon. 1 (bases 1 to 10677)			

FEATURES	source
JOURNAL TITLE	Roest-Crollius, H., Tallon, O., Dasilva, C., Fizames, C., Fisher, C., Weissenbach, J., Billaud, A., Queller, F., Saurin, W., Bernot, A. and
JOURNAL REFERENCE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL AUTHORS	Unpublished
TITLE	2 (bases 1 to 1067)
JOURNAL REFERENCE	Roest-Crollius, H., Tallon, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Winkler, P., Brotlier, P., Queller, F., Saurin, W. and Weissenbach, J.
JOURNAL AUTHORS	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
TITLE	Unpublished
JOURNAL REFERENCE	3 (bases 1 to 1067)
JOURNAL AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL REFERENCE	Submitted (12-Apr-2000) to the EMBL/Genbank/DBJ databases
JOURNAL AUTHORS	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon .
FEATURES	Location/Qualifiers
source	1..1067
	/organism="Tetraodon nigroviridis"
	/db_xref="taxon:99883"
	/clone_1=1/6609
	/clone_1lb="G"
	/note="Genoscope sequence ID : COAG176AD05SP1-end : PUC-Or1"
BASE COUNT	250 a 249 c 239 g 328 t 1 others
ORIGIN	
Query Match	6.3% Score 87.6; DB 12; Length 1067;
Best Local Similarity	54.3% Pred. No.1,3e-11;
Matches 202: Conservative	0; Mismatches 164; Indels 6; Gaps 1;
87	agaattgagcagaagaaatgaccagcagaatgaagacatcctaagcatcc
626	acaaatgacgacaaatcggaacccgattaccagagatgaaataattctgaacgagatcaga
147	tcttgaaaaagaaatggtgaagaagacctcttcaaatgctgtagtgcagaagcgacaga
566	tgaggaggggtgacattctttaaagacctcttcaaaaacgacgaagatgctggtgcgaaca
207	aatcgttttggttgattgattcctagacagcatcagttgataga-----tattgatga
506	atgcttcgttccctggtggattttcagaaacgcacaaatrtatctctgagagactatgtgacc
261	taagtgcgcccccatgacgaagcgacacattgtgtgtaacaaacacagccattacaga
446	cgacatmgagctctctgacagggcccttgcctggtggcggttaattatgacacgttcacaga
321	agatgatgttagaagaatcagaatccttgcgaagaagcgacgaagggaggaatccttata
386	tgacacactgacaaacattctctcctggtggcctctgcctcaaaaggaaacacacactgagaa
381	aactgcacgattgagaaatgaatgaatcattcgtgtatcatatcacagactgcccattt
326	atmgcgaagatttgcccttggatttcagacactgtgattcatgtgacagacatcccttcatt
441	tattcttgcaaa 452
266	cctcagtgagaa 255
RESULT 12	
BF650306	586 bp mRNA linear EST 20-DEC-2000
LOCUS	NP087D11ECP1092 Elicited cell culture Medicago truncatula cDNA
DEFINITION	clone NF087D11EC 5', mRNA sequence.
ACCESSION	BF650306
VERSION	BF650306.1 GI:11915436

KEYWORDS	EST.
SOURCE	barrel medic.
ORGANISM	Medicago truncatula
	Eumariotia, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE	1 (bases 1 to 586)
AUTHORS	Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores, H.R., Inman, J.T., Meller, J.W. and May, G.D.
TITLE	Expressed Sequence Tags from the Samuel Roberts Noble Foundation - Center for Medicago Genomics Research
JOURNAL	Unpublished (2000)
COMMENT	Contact: Dixon RA Plant Biology Division The Samuel Roberts Noble Foundation 2510 Sam Noble Parkway, Ardmore, OK 73402, USA Tel: 580 221 7302 Fax: 580 221 7380 Email: radixon@noble.org Insert length: 586 std Error: 0.00 Plate: 087 row: D column: 11 Seq primer: TCACACAGGAACGCTATGCAC.
FEATURES	Location/Qualifiers

FEATURES	SOURCE
location/Qualifiers	1. 586
/organism="Medicago truncatula"	/db_xref="taxon:3880"
/clone="NF087D11EC"	/clone_lib="Elcited cell culture"
/tissue_type="Cell cultures derived from root tissues"	/dev_stage="Cell suspensions were subcultured every 14 days. Cells were induced six days after subculture"
/note="Vector: lambda zap; Cells were induced with yeast cell wall extracts equivalent to 50ug/ml glucose in the final concentration. Samples were taken at 0.5', 1', 12 and 24 hours after induction. Equal amounts of RNA from each time point were pooled and used for mRNA isolation."	135 a 151 c 127 g 171 t 2 others

Query Match	6.28	Score 87.2	DB 10	Length 586
Best Local Similarity	51.28	Pred. No. 1.4e-11		
Matches 228, Conservative	0	Mismatches 214	Indels 3	Gaps 1

QY 74 caacattgagcagcaaatattggcgagaagaaattgacgcagcaattgaagcatcc 133
 Db 76 CAATTTCTTCGAGATTTTCGGCCAAACAGTGCACCTAATCTGTGCATCCCGAAGTTC 135
 QY 134 ttaatgcatalctctctgaaagaagaatgttgaagaagctcttcaaaatgctgatg 193
 Db 136 TCTTAACTATCCCGAAGGAACACCGGTTCTCAAGAGATTTATCGAAGGCGGATGAG 195
 QY 194 caaaggcgacagaataatcgttttctgtttgatacctagacagcatccagttgataat 253
 Db 196 CCGGTGTCACACCGCTTCTCTCTCTGTGACGCGCGCTCACACGCGCCCTGTAGTGTTC 255
 QY 254 ttgatgataagtgggcccccatcgacaggcgacgaacttggtgtacaacaacacgaat 313
 Db 256 TGTCCGATTTCTGTATTCGACAGTGGCAAGACCTCGCGCTTTTGGCTTATATGATGCTGTT 315
 QY 314 ttacagaagaatgattagagaatttcagaactcttgcgaagaagcgacgaagaaggaaatc 373
 Db 316 TCTGTGAAGAGATTTTGTAGTATTTGGAAGATCGGTGGGGAGTGTAGATGAGACAG 375
 QY 374 ctataaaaactggacagatggaatagaattcaattctgtgtatcatatcacagactgcc 433
 Db 376 CTTCCAAAACCGGTGCAATTCGGGGTTGGCTTCACCTCAGTGCATTTMAACAGATCTTC 435
 QY 434 catctttatcttcggcaatgacatcccggtgatttttgatccatgatgcagatgtgcc 493
 Db 436 CTTCAATTGCTGAGTGGCAAAATCTTGGTAT--TGTTTGACCTTCAGAGGTGTTTATCTTC 492

Qy	494	caggcgccatccatctagtc	ccg	518
Db	493	CAAGGTTCTGCANCA	AATCCTGG	517

RESULT 13
AZ136287

LOCUS	918 bp	DNA	linear	GSS 28-AUG-2000
DEFINITION	AZ136287			
	SP_0169_B2_B09-SP6E	Strongylocentrotus purpuratus		
	urc1in, sperm genomic BAC library	Strongylocentrotus purpuratus		
	genomic clone Plate=169	Col=18 Row=D, DNA sequence.		

ACCESSION	AZ136287
VERSION	AZ136287.1
KEYWORDS	GI:8288190 GSS.

SOURCE	ORGANISM
Strongylocentrotus purpuratus.	Strongylocentrotus purpuratus

Eukaryota; Metazoa; Echinodermata; Eleutherozoa
Echinoidea; Euechinoidea; Echinacea; Echinoida;
Echinoidea; Euechinoidea; Echinacea; Echinoida;

Strongylocentrotidae; Strongylocentrotus.
1 (bases 1 to 918)

AUTHORS
Cameron, R.A., Malhairs, G., Kasi, J.P., Martinez, P., Biondi, T.R., Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray, C. & Fennell, T.
Fennell, T., Cameron, R.A., Kasi, J.P., Martinez, P., Biondi, T.R., Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray, C. & Fennell, T.

HOOD, L.
G.A., ELEMENSU, C.A., DEMADEU, H., BILLEN, R.O., DAVIDSU, E.N. and
TITIE
A sea urchin genome project: Sequence scan virtual map and

JOURNAL
of the
Royal
Society
of
Medicine
1997
90
17
9514-9518
(2000)
Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
additional resources
sea urchin genome project, sequence scan, virtual map

MEDLINE 20402566
COMMENT Contact: Cameron, RA, Davidson, EH, Hood, L

Division of Biology 156-29
California Institute of Technology

Pasadena California 91125, USA
Tel: (626) 395-8421

Fax: (626) 793-3047
Email: acameron@caltech.edu

Plate: 169 row: D column: 18
Seq primer: SP6

Class: BAC ends
High quality sequence stop: 918.

```

FEATURES
  location/Qualifiers
    1. .918
      /organism="Stronau"]occurance"

```

```

/organism=strongylocentrotus [
/db_xref="taxon:7668"
/collection=plate169 Col=18 Row=D"

```

clone_11b="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"

```

/Note="Organ: sperm; Vector: BACe3.6; BAC clones in E-coli
DH10B"

```

BASE COUNT	262 a	207 c	214 g	235 t
ORIGIN				

Query Match	6.1%;	Score 85;	DB 12;	Length 918;
Best Local Similarity	54.0%;	Pred. No. 5.7e-11;		

Matches 222; Conservative 0; Mismatches 180; Indels 9; Gaps

84 cacagaatttgggcagaanaaatttgaccagcagaattgaagcattccttaatgcatc 143

DB 80 CACACCAGCTGGGCGACGACGACCCCTTGACTAAGCGCTTGGGAAATATCTCCAGAACCA 139

[illegible][illegible]

200 GGAATTCCTGCTTTCATTCGACATGAGAGAAAGCAACCTTTAGAGAAACCAACAGATCTTCGGA 259

258 tcaataacttgcgcccatttcacaaagccacgaccttttatgtatcacacaaacccacatttac 317

Db 260 TCCGGGATGAAATCTTGTACGGGACACGCCCTTTGGGTATACACGATGCCGATTTCAC 319

318 agaagatcatgttgaaggaattcgaatccttgaaaaagcgcgaaagaagggaatccctta 377

Db 320 AGACCAAGACTTGTGAGATATTTCTTCTTGAGGGGGGTACCAAGAAAGACGCTGA 379

Qy 378 taaactgcagatgataatgaatgaatcattcgtgtatcatatcacagactcccatc 437

Db 380 AAAGATTTGGGAGCTTTGGGACCGGTTTCAACTCGGTGTACCGCATTAATCTGTCCAA 439

Qy 438 ttattatcgcagatgacatccctgtgtattttgatccctatgcagata 488

Db 440 TTGTGTAGTAGGACTACATGAGT---TTTTCGACCCCTCATACACACA 487

RESULT 14

AZ185191 807 bp DNA linear GSS 30-AUG-2000

LOCUS SP_1004_B2.D05.T7A Strongylocentrotus purpuratus, purple sea urchin

DEFINITION , sperm genomic BAC library Strongylocentrotus purpuratus genomic

clone plate=1004 Col=10 Row=H, DNA sequence.

ACCESSION AZ185191

VERSION AZ185191.1 GI:8357566

KEYWORDS GSS.

SOURCE Strongylocentrotus purpuratus.

ORGANISM Strongylocentrotus purpuratus

Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;

Echinoidea; Euechinoidea; Echinacea; Echinoidea;

Strongylocentrotidae; Strongylocentrotus.

REFERENCE 1 (bases 1 to 807)

Camerton, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R.,

Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray

, G.A., Etlensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and

Hood, L.

A sea urchin genome project: Sequence scan, virtual map, and

additional resources

Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)

20402566

CONTACT: Camerton, RA, Davidson, EH, Hood, L

Division of Biology 156-29

California Institute of Technology

Pasadena California 91125, USA

Tel: (626) 395-8421

Fax: (626) 793-3047

Email: acameron@caltech.edu

Plate: 1004 row: H column: 10

Seq primer: T7

Class: BAC ends

High quality sequence stop: 807.

Location/Qualifiers

1. 807

/organism="Strongylocentrotus purpuratus"

/db_xref="taxon:7668"

/clone="Plate=1004 Col=10 Row=H"

/clone_lib="Strongylocentrotus purpuratus, purple sea

urchin, sperm genomic BAC library"

/note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli

DH10B"

BASE COUNT 245 a 182 c 186 g 192 t 2 others

ORIGIN

Query Match 6.0%; Score 84; DB 12; Length 807;

Best Local Similarity 54.2%; Pred. No. 1e-10;

Matches 218; Conservative 0; Mismatches 175; Indels 9; Gaps 2;

Qy 93 tgggcagaagaataatgaccagagaatgaagacatccatcattatcctctga 152

Db 35 TGGGCACGATGAACCTTGTACTATGCGCTTGGAATATCTCAAGAACACTATGTGGA 94

Qy 153 aaagaatgttgaagaagcttctcaaatgctgataatgacagagcagagaatctg 212

Db 95 CACGCGCATGTTAGTAAATGATACAGAAATGACAGAGATCAGAGCTCAGGAAGTTCG 154

Qy 213 ttctgtttgatcctgagc-----agcatcaggttgatagaatatttgatgaagtg 266

Db 155 CTTTCTCATGTGACATGAGAGAAGAACGAGAAAGAAACCAAGATTATTCATCCGGAA 214

Qy 267 gcccacattgcagaagccagacatttgtgtacacaacagacatttacagaatga 326

Db 215 GAAGCTCTGTGAGGACACGACCTTTGGGTATACAAACATCTGTATTCACAGACAGA 274

Qy 327 tgttagagaatgaaatccttggaaaaagcagaagaagggaatcctttaaactcg 386

Db 275 CTTTAAGAAATATTCCTCGTCTTGGAGGCGGTACCAAGAAAGACGCTGAAAGATCGG 334

Qy 387 acagatgaaatagatcaatcctcgtgtatcatatcacagactgccatcttattc 446

Db 335 GAAGTTTGGACCGGTTTCAACTCGGTCTACCGCATTAATCTGTTCAGATTGGTTAG 394

Qy 447 tggcaatgacatcctgtgtattttgatccctatgcagata 488

Db 395 TAGGACTACATGACAGT---TTTTCGACCCCTCATACACACA 433

RESULT 15

AZ182120 431 bp DNA linear GSS 30-AUG-2000

LOCUS SP_0188_A1.A09.SP6E Strongylocentrotus purpuratus, purple sea

DEFINITION urchin, sperm genomic BAC library Strongylocentrotus purpuratus

genomic clone plate=188 Col=17 Row=A, DNA sequence.

ACCESSION AZ182120

VERSION AZ182120.1 GI:8354495

KEYWORDS GSS.

SOURCE Strongylocentrotus purpuratus.

ORGANISM Strongylocentrotus purpuratus

Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinoidea;

Echinoidea; Euechinoidea; Echinacea; Echinoidea;

Strongylocentrotidae; Strongylocentrotus.

REFERENCE 1 (bases 1 to 431)

Camerton, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R.,

Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray

, G.A., Etlensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and

Hood, L.

A sea urchin genome project: Sequence scan, virtual map, and

additional resources

Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)

20402566

CONTACT: Camerton, RA, Davidson, EH, Hood, L

Division of Biology 156-29

California Institute of Technology

Pasadena California 91125, USA

Tel: (626) 395-8421

Fax: (626) 793-3047

Email: acameron@caltech.edu

Plate: 188 row: A column: 17

Seq primer: SP6

Class: BAC ends

High quality sequence stop: 431.

Location/Qualifiers

1. 431

/organism="Strongylocentrotus purpuratus"

/db_xref="taxon:7668"

/clone="Plate=188 Col=17 Row=A"

/clone_lib="Strongylocentrotus purpuratus, purple sea

urchin, sperm genomic BAC library"

/note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli

DH10B"

BASE COUNT 126 a 93 c 105 g 105 t 2 others

ORIGIN

Query Match 5.7%; Score 80; DB 12; Length 431;

Best Local Similarity 51.8%; Pred. No. 9.2e-10;

Matches 207; Conservative 0; Mismatches 187; Indels 6; Gaps 1;

Qy 84 cagagaattggcagaagaataatgaccagcagaatgaagacatccatcattatc 143

Db 9 CACACACCTGGCACAATGAACCTTGACTATGCGCTTGTGGAATATTCACAGACAA 68

```
QY 144 tcctctgaaaggaatgttgaaagagcttctcaaaatgtgatgtgcaagcgac 203
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 69 CTTTGTGGACACCGGTATCGTTAGTGAATGATACAGATCGGAGATGCGAGACTCA 128
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 204 agaaatctgttltgtgttgatcctagac-----agcaccagltgataaataattga 257
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 129 GGAAGTTGCTTCTCATTTGACATGAGAAAGCAAGAAAGAAACCAACAAATTGTTGCA 188
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 258 tgataagtgagcccatgtgcaaggccagcaattgtgtgtaacaacaaccagccattac 317
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 189 TCCGGGATGAAGTCTTCTCAGGGGCCAGCCCTTGGGTATACACGATGCCGTATTTCAC 248
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 318 agaagatgagtctagaaggaattcagaatcttgaaagcaagcaagaggaatccctta 377
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 249 AGACCAAGACTTTGAGATATTTCCTCTAGAGGGCGGTACCAAGAAAGAGACGCTGA 308
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 378 taaacttgacagatggaatagaaatcgaatctgtgtatcatatcacagaactgcccac 437
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 309 AAAGATTGNGAAGTTTGGACCGGTTTCACCTCGGTACCGCATTAACCTGATGTTCCAAG 368
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 438 tttaattctgcaatgacacccctgtgtattttgatcct 477
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 369 TTNGGTATAGGGACTACATGACAGATTTCCTACCCCTCAT 408
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: May 22, 2002, 05:31:27
Job time: 4087 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 07:30:25 ; Search time 2968.03 Seconds
(without alignments)
6370.971 Million cell updates/sec

Title: US-09-693-205-7_COPY_5300_6700

Perfect score: 1401

Sequence: 1 gtagcagcagtaaacatagagc.....aaagacacttaagaagt 1401

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 674847542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estcin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gD_est1:*
10: gD_est2:*
11: gD_htc:*
12: gD_gss:*
13: em_gss_hum:*
14: em_gss_iny:*
15: em_gss_pln:*
16: em_gss_vtl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	656	46.8	1083	BM476887	BM476887 AGENCOURT
2	312	22.3	376	AL596565	AL596565 DKF2P451L
3	205	14.6	1010	BM472954	BM472954 AGENCOURT
4	201	14.3	209	AA776169	AA776169 aeo8c02.s
5	169	12.1	795	R17106	R17106 EST20108 C1
6	140	10.0	148	AA776670	AA776670 aeo80902.s
7	140	10.0	238	AA897178	AA897178 am09e08.s
8	7	1.6	665	BM157201	BM157201 ENTS105TF
9	21	1.5	247	AV214110	AV214110 AV214110
10	21	1.5	605	BM034007	BM034007 RCTI-24-2
11	21	1.5	517	BM034007	BM034007 RCTI-24-2
12	21	1.5	605	BM034007	BM034007 RCTI-24-2
13	20	1.4	758	B70904	B70904 CIT-HSP-206
14	20	1.4	308	BM624213	BM624213 CEMB-140
15	20	1.4	320	AI705955	AI705955 UI-R-ACO-
16	20	1.4	331	AI579360	AI579360 UI-R-AGO-
17	20	1.4	355	BM295731	BM295731 UI-R-DK0-
				AM435291	AM435291 UI-R-BD0P

18	20	1.4	374	9	AM520724	AM520724 UI-R-BJ0P
19	20	1.4	383	9	BE101372	BE101372 UI-R-BJ1-
20	20	1.4	384	9	AM433545	AM433545 UI-R-BJ0P
21	20	1.4	418	12	AQ153679	AQ153679 HS-224_B
22	20	1.4	461	9	AI043792	AI043792 UI-R-CO-j
23	20	1.4	464	9	AI704578	AI704578 UI-R-ACO-
24	20	1.4	477	9	AA98882	AA98882 UI-R-CO-h
25	20	1.4	477	9	AI710607	AI710607 UI-R-AC1-
26	20	1.4	517	10	BM287243	BM287243 527562 MA
27	20	1.4	522	10	BE912359	BE912359 601664865
28	20	1.4	534	10	BE12760	BE12760 RH41717.5
29	20	1.4	535	10	BM257525	BM257525 521174 MA
30	20	1.4	541	10	BE668948	BE668948 159386 MA
31	20	1.4	543	9	BE103356	BE103356 UI-R-BX0-
32	20	1.4	546	12	AQ534831	AQ534831 RPTI-11-3
33	20	1.4	563	10	BI290791	BI290791 UI-R-DK0-
34	20	1.4	568	9	AL638993	AL638993 AL638993
35	20	1.4	571	12	AQ485516	AQ485516 RPTI-11-2
36	20	1.4	616	12	AZ963183	AZ963183 2M0232F17
37	20	1.4	621	12	AQ446973	AQ446973 mgXD00021
38	20	1.4	652	10	BJ093045	BJ093045 BJ093045
39	20	1.4	693	9	BB365802	BB365802 BB365802
40	20	1.4	695	9	AL636882	AL636882 AL636882
41	20	1.4	701	9	AL631304	AL631304 AL631304
42	20	1.4	705	12	AQ255211	AQ255211 mgxb0009B
43	20	1.4	930	10	BF163742	BF163742 601768844
44	20	1.4	931	12	AZ204179	AZ204179 SP-0097_A
45	19	1.4	106	12	AQ283816	AQ283816 RPTI11-78

ALIGNMENTS

RESULT 1
LOCUS BM476887
DEFINITION BM476887 1083 bp mRNA linear EST 05-FEB-2002
5' mRNA sequence.
ACCESSION BM476887
VERSION BM476887.1 GI:18525929
KEYWORDS EST.

SOURCE

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1083)
NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMN)

DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC Clone Distribution Information can be

found through the I.M.A.G.E. Consortium/LMN at:
http://image.llnl.gov

Plate: LAM12275 row: j column: 18

High quality sequence stop: 696.

Location/Qualifiers

1. 1083

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:555441"

/clone_lib="NIH-MGC_71"

/tissue_type="leiomysarcoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: uterus; Vector: PCMV-SPORE6; Site:1; NotI;

Site:2; SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 2.1 kb.

334 a 219 c 212 g 318 t

BASE COUNT
ORIGIN

Query Match 46.8%; Score 656; DB 10; Length 1083;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 656; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 640 gcagaatgcaaaagtttggaaatttcgtcttccagatcagacagaatgtccag 699
 |||||||
 DB 13 GCAGAAATGGCAAAAGTTTGGAAATTCGCTCTCCAGCATGACAGAAATGGTCCAG 72
 QY 700 aacttttgcaaacctgcctcagatcgggagcagaacttcaatgtttcctacacatg 759
 |||||||
 DB 73 AATCTTTGGCAAACTGCGCTCAGATGGGCGACAACTTCTAATGTTCTTAATCACATG 132
 QY 760 gaaaaatttcattctgtgaataatagataagatcgtgaagctcctaaatgtgtctatca 819
 |||||||
 DB 133 GAAAAATTTCTATTGTAATAGATAGAGTACTGAGCTGAAATGTCTGTATTC 192
 QY 820 gtaaaaggcaaaatcacagatggagacagatgtgaaaggaaacatttcagctcgtga 879
 |||||||
 DB 193 GTAAGGGCAAAATCACAGATGAGACAGATTGAAAAGAAATTCATGATCATCTGTA 252
 QY 880 attgatgttacttaaaagagcagctcaagacataccagttcaacaataacctat 939
 |||||||
 DB 253 ATTGATAGCTGTACTAATAAAGAGCAGCTCAAAACATACAGTTCAACAAATACCTAT 312
 QY 940 actatgatactgagagactctgaagaatcttactacgtgtgctaatttgaatagatca 999
 |||||||
 DB 313 ACTATGATAGCTGTACTGAGTGAAGAAATCTTACTAGTGGCTAATTTGTAATAGATCA 372
 QY 1000 ggccttcaagatggagaaagtatctaaagtgtcatalatcagttcaagaaccaagat 1059
 |||||||
 DB 373 GGCCTTCAAGTATGAGAAAGTATCTAATAAGTGTCAATCAGCTCAACAAACCAAGAT 432
 QY 1060 attactcttccacacgtgtgtgagtagctgcctgcattactcacaactaaacaaacc 1119
 |||||||
 DB 433 ATTACTCTTTCCACAGCTGGTGGAGTAGCTGCCCTGCAATTACTCAACTATAAACC 492
 QY 1120 catagggcctctgtcttctgcctctcttcttgagagctgggtgcatttcattgaat 1179
 |||||||
 DB 493 CATAGGGCCTCTCTGTTTTCCTCTTCTTTGGAGACTGGGCTGCCATTTCAATGTAAT 552
 QY 1180 ggcacatttgacactgagttcagcagaagaagactgtggcgttgatgtaagggttgc 1239
 |||||||
 DB 553 GGCACACTTGGACTGAGTCCAGCCAGAAAGAACTGTGGCTGATGATAAGGAGTTGT 612
 QY 1240 gtccgaagtgaactgaataacagtttaatagcacagcatttaagtcctgcatacgt 1295
 |||||||
 DB 613 GTTCGAAGTACTGGAATACAGTTTAATGACACATTAAAGCTCCTGCATATGT 668

RESULT 2
 AL596565 376 bp mRNA linear EST 14-AUG-2001
 LOCUS DKEZ451L0110.F1.451 (synonym: hlcc1) spinal cord Homo sapiens cDNA
 DEFINITION clone DKEZ451L0110.5, mRNA sequence.
 ACCESSION AL596565
 VERSION AL596565.1 GI:15154261
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 376)
 AUTHORS Othenwaelder,B., Obermaier,B., Mewes,W., Mewes,H.W., Well,B. and Wiemann,S.
 TITLE (Othenwaelder,B., Obermaier,B., Mewes,H.W., Well,B. and Wiemann,S.)
 JOURNAL Unpublished (2001)
 COMMENT Contact: Othenwaelder B
 MIPs Am Kioferplatz 18a D-82152 Martinsried, Germany
 This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de; sequenced by Medigenomix (Martinsried/Germany) within the cDNA sequencing consortium of the German Genome Project. No sl sequence available.

This clone (DKFZ451L0110) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES
 source
 1.376
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="DKFZ451L0110"
 /clone_lib="451 (synonym: hlcc1) spinal cord"
 /tissue_type="human spinal cord"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI"

BASE COUNT 127 a 63 c 79 g 107 t

Query Match 22.3%; Score 312; DB 9; Length 376;
 Best Local Similarity 100.0%; Pred. No. 2.8e-152; Length 376;
 Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 550 aggacacagttctcagatgtcttgcattatctcgtggaacccatttaactgataat 609
 |||||||
 DB 1 AGGACACAGTTCCTCAGATGTTCTGATCTTATCTGGAACCCATTTAACTGATAAT 60
 QY 610 tgcacatgttccagatttctctcttcgtgaatgcagaatgtcagaatgttggaaatttcg 669
 |||||||
 DB 61 TGCAAAATGTTCAAGATTTCCTCTTCTGTAATGCAAAATGCGAAAGTTTGGAAATTCG 120
 QY 670 tctgtccagatcagacagaatgttccagaatcttcttgacaacacgcgtcagatgg 729
 |||||||
 DB 121 TCTGTTCCAGATCAGACAGATGTCAGATCTTTTGGCAAACTGCCCTCAGATGG 180
 QY 730 gcagaactctcaatgttcttcaatcacatggaataattctatttgaatagataag 789
 |||||||
 DB 181 GCAGAACTTCAATGTTCTTAATCACATGGAATAAATTTCTATTGTAATAGATAAG 240
 QY 790 agtctcagctcctcaatgtcgtgattcagtaaggcaaaatcacagatgtgagacaga 849
 |||||||
 DB 241 AGTACTGAGCTCAAAATGTCTGATTCAGTAAGGCAAAATCACAGATGAGACAGA 300
 QY 850 ttgaaaaggaaa 861
 |||||||
 DB 301 TTGAAAAGGAAA 312

RESULT 3
 BM472954 1010 bp mRNA linear EST 05-FEB-2002
 LOCUS BM472954 AGENCOURT_6466106 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5574588
 DEFINITION 5', mRNA sequence.
 ACCESSION BM472954
 VERSION BM472954.1 GI:18521996
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1010)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
 TITLE NIH-MGC National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs.femail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I M A G E Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov
Plate: LLAM12323 row: h column: 13
High quality sequence stop: 738.

Location/Qualifiers

1. 1010

FEATURES

source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:557458"
/clone_lib="NIH_MGC_88"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: small intestine; Vector: pCMV-SPORT6;
Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
oligo-dr primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 293 a 209 c 191 g 317 t
ORIGIN

Query Match

14.6%; Score 205; DB 10; Length 1010;

Best Local Similarity 99.5%; Pred. No. 3.8e-96;

Matches 375; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 920 cagttcaacaatacactactatgatactcgaagactctgaagaaacttactactgt 979

DB 20 CAGTTCAACAATAACCTATACATGATGACTGAGGACTCTGAAGCAAACTTACTACTGT 79

QY 980 ggttaatttgtaataagtcgagcttttcaagtagtgagaagatratcctaaagtgctat 1039

DB 80 GGTAAATTGTATATAGTTCAGGCTTTCAAGTATGAGAAAGATCTAAAGTGTCTATAT 139

QY 1040 cagctcaagaacaagatatactctcttccacagtggtgagtagctgcctgcactta 1099

DB 140 CAGCTCAACAAGAACAGATATACCTTTTCCACACTGGTGGAGTAGCTGCCGCACTTA 199

QY 1100 ctacaactat-aaaaacccatagggcctctgttttttgcctcttcttcttggagact 1158

DB 200 CTCACAACATTAATAAAAAACCCCATAGGCGCTTCGTTTTCCTCTTCTTGAGAGACT 259

QY 1159 gggctgcattctatgtgaatggcactcttgcactgtaaccagcaagaagactctgtg 1218

DB 260 GGGCTGCATTTTCATGGAATGGCCACTTTGCACCTGATTCAGCCACAAGAACCTTGG 319

QY 1219 cgtgatgataatgagtggtgtgttcgaagtgactggaataacagtttaatacgagacta 1278

DB 320 CGTGATGATTAATGAGTGTGTGTTCGAAAGTGACTGGAAATACAGTTTAATGACAGCACTTA 379

QY 1279 atagctctgcataatgt 1295

DB 380 ATAGCTCTGCATATATGT 396

RESULT 4

AA776169

LOCUS AA776169 209 bp mRNA linear EST 05-FEB-1998

DEFINITION ae80c02.s1 Stragene schizo brain S11 Homo sapiens cDNA clone

IMAGE:970466 3', mRNA sequence.

ACCESSION AA776169

VERSION AA776169.1 GI:2835503

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 209)

AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,

Kritzman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marita,M., Martin

J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,

White,Y., Wylie,T., Waterston,R. and Wilson,R.

TITLE WashU-NCI human EST Project

JOURNAL

Unpublished (1997)

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Possible reversed clone: polyT not found

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 206.

FEATURES

source

1. 209

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:970466"

/clone_lib="Stragene schizo brain S11"

/sex="male"

/tissue_type="schizophrenic brain S-11 frontal lobe"

/dev_stage="34 years old"

/lab_host="SOLR (kanamycin resistant)"

/note="Vector: Bluescript SK-; Site_1: EcoRI; Library

constructed from S-11 frontal lobe, male, 34 years old,

50% caucasian, 50% Aleutian. Schizophrenic suicide.

Random primed into EcoRI site of Zap II Vector. Mass

excised. Custom library. Avg insert length 1.4kb.

Material obtained by Johnston N., Torrey, E.F., Yolken R.,

and the Stanley Neuropathology Consortium - Analysis of

RNAs from the Brains of individuals with Psychiatric

Diseases (unpublished) Stanley Neuropathology Laboratory,

Johns Hopkins School of Medicine, Baltimore MD."

BASE COUNT 58 a 42 c 47 g 62 t
ORIGIN

Query Match

14.3%; Score 201; DB 9; Length 209;

Best Local Similarity 100.0%; Pred. No. 5.1e-94;

Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 501 caactccattagtcggagcagcatgttagagacttgatgcagatcttagacacagtt 560

DB 9 CACATCCATTATAGTCCGAGCAGCATGTTATAGATTTGGATGCACATTTTAGACACAGTT 68

QY 561 ctcaagatgttcggatcttatacttggaaccatttaactggataatgacacagtt 620

DB 69 CTGAGATGTTCTGATCTTTATCTGGACCCCATTTAACTGATTAATTGCACAAATGTT 128

QY 621 cagattccctcttcgttaatgcagaatgcaaaagtttcggaaatcttcgtcttcagc 680

DB 129 CAGATTTTCCTCTTCGTATATCAGAAATGCGAAAGTTTCGAAATTTTCGTTCCAGC 188

QY 681 atcagacagaatggtccagaa 701

DB 189 ATCAGACAGATGTCAGAA 209

RESULT 5

R17106

LOCUS R17106 795 bp mRNA linear EST 12-JUN-1996

DEFINITION EST20108 Clontech adult human fat cell library HL1108A Homo sapiens

CDNA clone 20108, mRNA sequence.

ACCESSION R17106

VERSION R17106.1 GI:770716

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 795)

AUTHORS Boulland,F.

TITLE Study of expressed sequences tags in adipose tissue 1995

JOURNAL Unpublished (1995)

COMMENT Contact: Frederic Boullaud
Centre de Recherche sur l'Endocrinologie moléculaire et le
Développement
CNRS
9, Rue Jules Hetzel, Meudon Bellevue, 92190 France
Tel: 33 1 45 07 52 87
Fax: 33 1 45 07 58 90
Email: boullaud@infobiogen.fr
Southern human DNA EcoRI single band 2.9 Kb.

FEATURES
source
Location/Qualifiers
1. 795
/organism="Homo sapiens"
/strain="Caucasian"
/db_xref="taxon:9606"
/clone="20108"
/note="Vector: lambda gt10; Site 1: EcoRI; dev-stage=adult
; tissue-type=adipose tissue; lab_host=Bacteriophage
lambda; first strand priming with random and poly-d(T)
oligonucleotides."

BASE COUNT 243 a 150 c 166 g 216 t 20 others

ORIGIN

Query Match 12.1%; Score 169; DB 10; Length 795;
Best Local Similarity 99.3%; Pred. No. 3e-77;
Matches 269; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 349 ggaagagccggaagaggaagaaacccctataaacctgagcagatggaaatgaattcaat 408
|||||
Db 21 GGAAGAGCCGGAAGAGGGAAGAACTTATTAAGATGAGATGGAATGAGATTCAT 80
|||||

QY 409 tccgtgatcatatcacagagctgccatcttattctctggaatgacatccgtgatt 468
|||||

Db 81 TCTGTGATATATTCACAGACTGCCCATCTTTATTCTGCAATGACATCTGTGATT 140
|||||

QY 469 ttgatctccatccagatatgacacagggccacatcatagtcctccgacgcatgtt 528
|||||

Db 141 TTGATATCTCATGCAAGATATGCACAGGGCCACATCATTAAGTCCCGAGCATGT 200
|||||

QY 529 aagagattgagatcagatttagaacacagtttcagatgttctgtcttctcggga 588
|||||

Db 201 AGGATTTGGATGAGATTTTAGACACAGNTCTCAGATGTCTGATCTTTATCTGGA 260
|||||

QY 589 acccatcttaactgataatgcacatgt 619
|||||

Db 261 ACCCATTTAACTGATATTCACACATGT 291
|||||

RESULT 6
AA776670 148 bp mRNA linear EST 05-FEB-1998
LOCUS ae80602.s1 Stratagene schizo brain S11 Homo sapiens cDNA clone
DEFINITION IMAGE:970514.3', mRNA sequence.
ACCESSION AA776670
VERSION AA776670.1 GI:2836004
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 148)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisels, G., Jost, S.,
Kritman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Maita, M., Martin
J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
White, Y., Wylie, T., Waterston, R., and Wilson, R.
WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Wilton RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

EMAIL: est@watson.wustl.edu
This clone is available royally-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Possible reversed clone: polyT not found
Seg primer: -40m3 fwd. ET from Amersham
High quality sequence stop: 131.
Location/Qualifiers
1. 148
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:970514"
/clone_lib="Stratagene schizo brain S11"
/sex="male"
/tissue_type="schizophrenic brain S-11 frontal lobe"
/dev_stage="34 years old"
/note="Vector: Bluescript SK-; Site 1: EcoRI; library
constructed from S-11 frontal lobe, male, 34 years old,
50% caucasian, 50% Aleutian. Schizophrenic suicide.
Random primed into EcoRI site of ZAP II Vector. Mass
excised. Custom library. Avg insert length 1.4kb.
Material obtained by Johnston N., Torrey, E.F., Yolken R.,
and the Stanley Neuropathology Consortium - Analysis of
RNAs from the Brains of individuals with Psychiatric
Diseases (Unpublished) Stanley Neuropathology Laboratory,
Johns Hopkins School of Medicine, Baltimore MD."

BASE COUNT 37 a 30 c 33 g 48 t

ORIGIN

Query Match 10.0%; Score 140; DB 9; Length 148;
Best Local Similarity 100.0%; Pred. No. 5.3e-62;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 501 cacatccatagatcccgagcagatgttagagattgagatgacagatttagaacacagt 560
|||||

Db 9 CACATCCATTATGTCGCGAGCATGTGTTAGATTTGATGATGATTTAGACACAGT 68
|||||

QY 561 ctcaagatgtctgagatcttatctgggaaccccttaactgataatgcacatgtt 620
|||||

Db 69 CTCAGATGTTCTGAGATCTTATCTGGGAACCATTTTAACGTGATATTGCCATGTT 128
|||||

QY 621 cagatttcctcttcgtaattg 640
|||||

Db 129 CAGATTTCCCTCTGTGTAATG 148
|||||

RESULT 7
AA897178/c 238 bp mRNA linear EST 04-JAN-1999
LOCUS am09e08.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:1466342.3', mRNA sequence.
ACCESSION AA897178
VERSION AA897178.1 GI:3033798
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 238)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncigap.
TITLE Tumor Gene Index
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapds-rc@mail.nih.gov
This clone is available royally-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 847 Std Error: 0.00
Seg primer: -40m3 fwd. ET from Amersham
High quality sequence stop: 132.
Location/Qualifiers
1. 238

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1466342"
/clone_lbp="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/notes="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NDHL19W, testis NHT, and B-cell
NCI-CGAP_GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT      71 a      53 c      45 g      69 t
ORIGIN

Query Match      10.0%; Score 140; DB 9; Length 238;
Best Local Similarity 100.0%; Pred. No. 5.2e-62;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1156 actgggcgtccatttcgtgaatgagccacttgcactggaattcaagcagaagaactgt 1215
    |||||||
DB 238 ACTGGGCTGCATTTCAATGTAATGGCCACTTTCACATGATGACCCAGAACGACTGT 179

QY 1216 tggagtgatgaataatgagtgctgttcgaagtgactggaataacagttaatacagcaga 1275
    |||||||
DB 178 TGGCGTGATGATATGAGTGTGCTTCGAACTGACATGGAATACAGTTTATATGACAGCA 119

QY 1276 ttaatagcctcctgcatatgt 1295
    |||||||
DB 118 TTAAATAGCTCCTGCATATGT 99

RESULT 8
BH157201/c 665 bp DNA linear GSS 24-SEP-2001
LOCUS EMTSL050TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic DNA sequence.
ACCESSION BH157201
VERSION BH157201.1 GI:15730639
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 665)
AUTHORS Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HM1:IMSS sheared DNA library (2001)
COMMENT Unpublished (2001)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b1loftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
DNA library
Seq primer: M13-Forward
Class: Shotgun
High quality sequence start: 18
High quality sequence stop: 569.
Location/Qualifiers
1..665
/organism="Entamoeba histolytica"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lbp="Entamoeba histolytica Sheared DNA"
FEATURES
SOURCE

```

```

/notes="Vector: pHOSt; Site.1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + 1 method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaubin and B.
Barrell, Oxford University Press, 1999)."
BASE COUNT      237 a      125 c      82 g      221 t
ORIGIN

Query Match      1.6%; Score 22; DB 12; Length 665;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 gaatcgtgttggttgatc 226
    |||||||
DB 28 GAATCTGTTTGTCTTGATC 7

RESULT 9
AV214110 247 bp mRNA linear EST 30-OCT-1999
LOCUS AV214110 RIKEN full-length enriched, ES cells Mus musculus cDNA
DEFINITION clone 2410133H03.3' similar to AF015309 Mus musculus nucleolar
protein (MSP58) mRNA, mRNA sequence.
ACCESSION AV214110
VERSION AV214110.1 GI:6154956
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 247)
AUTHORS Konno,H., Alzawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,T.,
Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugihara,Y., Suzuki,H.,
Suzuki,H., Takahashi,F., Tateo,M., Tomioka,N., Tsunoda,Y.,
Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
Yoshiki,A., Yoshino,M., Yamatsus,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al. 1999)
Unpublished (1999)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
Matsura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki
,Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh,M., Katsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,

```


OY 853 aaagaacaattcatgca 873
 Db 504 AAAGGAACATTCATGCA 524

RESULT 12
 B70904 758 bp DNA linear GSS 21-JUN-1998
 LOCUS CTT-HSP-2063E16.TF CTT-HSP Homo sapiens genomic clone 2063E16, DNA
 DEFINITION sequence.
 ACCESSION B70904
 VERSION B70904.1 GI:2710128
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 758)
 Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
 K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.
 and Venter,J.C.
 Use of a random BAC End Sequence Database for Sequence-Ready Map
 Building
 Unpublished (1997)
 Other-GSS: CTT-HSP-2063E16.TR
 CONTACT: Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadams@tigr.org
 Clones are available from Research Genetics (info@resgen.com). BAC
 end search page:
 http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
 Seq primer: M13-21
 Class: BAC ends.

FEATURES
 source Location/Qualifiers
 1..758
 /organism="Homo sapiens"
 /db_xref="GDB:7061491"
 /db_xref="taxon:9606"
 /clone="2063E16"
 /clone_1lb="CTT-HSP"
 /sex="Male"
 /cell_type="Sperm"
 /note="Vector: pBelobAC11; site_1: HindIII; site_2:
 HindIII"

BASE COUNT 259 a 133 c 123 g 243 t

ORIGIN

Query Match 1.5%; Score 21; DB 12; Length 758;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1332 ggttcgtatccacattatca 1352
 Db 736 GGTCGTGATCCACATATATCA 756

RESULT 13
 B624213 227 bp mRNA linear EST 01-JUL-2001
 LOCUS B624213
 DEFINITION CUEMB-140 CUEMB Crassostrea virginica cDNA, mRNA sequence.
 ACCESSION B624213
 VERSION B624213.1 GI:14580643
 KEYWORDS EST.
 SOURCE eastern oyster.
 ORGANISM Crassostrea virginica
 Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Ostreoida;
 Ostreidae; Ostreidae; Crassostrea.

REFERENCE 1 (bases 1 to 227)
 AUTHORS Gross,P.S. and Bartlett,T.C.
 TITLE Crassostrea virginica Embryo EST Library (CUEMB)
 JOURNAL Unpublished (2001)
 COMMENT Contact: Jenny MJ
 Marine Biomedicine and Environmental Sciences
 Medical University of South Carolina
 221 Port Johnson Road, Charleston, SC 29412, USA
 Tel: 843 876 5058
 Fax: 843 762 5350
 Email: jenny@emusc.edu.

FEATURES
 source Location/Qualifiers
 1..227
 /organism="Crassostrea virginica"
 /db_xref="taxon:9565"
 /clone_1lb="CUEMB"
 /tissue_type="Embryo, D-veliger larvae"
 /note="Vector: pTriblex2; site_1: Sfi I; site_2: Sfi I"

BASE COUNT 71 a 40 c 72 t

ORIGIN

Query Match 1.4%; Score 20; DB 10; Length 227;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1013 tggagaaagtatctaaagt 1032
 Db 48 TGGAGAAAGTATCTAAAGT 67

RESULT 14
 A1705955 308 bp mRNA linear EST 03-JUN-1999
 LOCUS UI-R-AC0-Y1-b-07-0-UI.s1 UI-R-AC0 Rattus norvegicus cDNA clone
 DEFINITION UI-R-AC0-Y1-b-07-0-UI 3', mRNA sequence.
 ACCESSION A1705955
 VERSION A1705955.1 GI:4993855
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 308)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 CONTACT: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Oligo-dT track not found. Not 1 site shown in beginning of sequence
 is likely internal to the message. cDNA Library Preparation: M.B.
 Soares lab clone distribution: clones will be available through
 Research Genetics (www.resgen.com) The following repetitive
 elements were found in this cDNA sequence: 263-301, >(CAG
)n#Simple-repeat
 Seq primer: M13 Forward
 POLYA-No.

FEATURES
 source Location/Qualifiers
 1..308
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-AC0-Y1-b-07-0-UI"
 /clone_1lb="UI-R-AC0"
 /dev_stage="adult"

/lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-AGO
 library is a non-normalized library constructed from 16.5
 dpc rat atriocentricular (AV) canal. The tag is a string
 of 5 nucleotides present between the Not I site and the
 oligo-dt track. The library was constructed as described
 by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
 , 1996. Tissue provided by Jim Lin, Department of Biology,
 University of Iowa.
 TAG_SPO=None found"
 BASE COUNT 46 a 77 c 78 g 107 t
 ORIGIN

Query Match 1.4%; Score 20; DB 9; Length 308;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1144 cttctcttggaactggct 1163
 ||||||||||||||||
 Db 7 CTTCTTTGGAGACTGGGCT 26

RESULT 15
 AI579360 320 bp mRNA linear EST 05-APR-1999
 LOCUS UI-R-AGO-wu-h-08-0-UI.s1 UI-R-AGO Rattus norvegicus cDNA clone
 DEFINITION
 ACCESSION AI579360
 VERSION AI579360.1 GI:4563736
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 320)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Oligo-dt track not found. Not I site shown in beginning of sequence
 is likely internal to the message. cDNA library preparation: M.B.
 Soares Lab Clone distribution: clones will be available through
 Research Genetics (www.resgen.com) The following repetitive
 elements were found in this cDNA sequence: 264-317, >(CAG
)n\$imple_repeat
 Seq primer: M13 Forward.
 Location/Qualifiers
 1..320

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 COMMENT

FEATURES

source
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-AGO-wu-h-08-0-UI"
 /clone.lib="UI-R-AGO"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-AGO
 library is a non-normalized library constructed from 13
 dpc rat ventricle. The tag is a string of 6 nucleotides
 present between the Not I site and the oligo-dt track.
 The library was constructed as described by Bonaldo,
 Lennon and Soares, Genome Research 6: 791-806, 1996.

Tissue provided by Jim Lin, Department of Biology,
 University of Iowa."
 BASE COUNT 47 a 80 c 80 g 113 t
 ORIGIN

Query Match 1.4%; Score 20; DB 9; Length 320;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1144 cttctcttggaactggct 1163
 ||||||||||||||||
 Db 7 CTTCTTTGGAGACTGGGCT 26

Search completed: May 22, 2002, 07:30:35
 Job time: 10805 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 08:30:10 ; Search time 3328.52 seconds
(without alignments)
8808.142 Million cell updates/sec

Title: US-09-693-205-7_COPY_5300_6700
Perfect score: 1401
Sequence: 1 gtacgagtaaacactagagc.....aaagacacttaagaagt 1401

Scoring table:
OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenBank: 1: gb_ba: 2: gb_bt: 3: gb_in: 4: gb_om: 5: gb_ov: 6: gb_ph: 7: gb_pl: 8: gb_pr: 9: gb_ro: 10: gb_ro: 11: gb_sts: 12: gb_sy: 13: gb_un: 14: gb_vl: 15: gb_ba: 16: em_fun: 17: em_hum: 18: em_in: 19: em_mu: 20: em_om: 21: em_or: 22: em_ov: 23: em_pat: 24: em_ph: 25: em_pl: 26: em_ro: 27: em_sts: 28: em_un: 29: em_vl: 30: em_hg_hum: 31: em_hg_inv: 32: em_hg_other: 33: em_hgo_inv:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query No. Score Match Length DB ID Description

1	1295	92.4	12793	6	AX119931	AX119931 Sequence
2	1295	92.4	12793	9	AF193556	AF193556 Homo sapi
3	1295	92.4	92693	9	AL157766	AL157766 Homo sapi
4	923	65.9	99819	2	AC079761	AC079761 Homo sapi
5	53	3.8	11492	6	AX119933	AX119933 Sequence
6	53	3.8	11493	10	AF193557	AF193557 Mus muscu
7	23	1.6	189760	9	AC022025	AC022025 Homo sapi
8	22	1.6	163869	9	AL356967	AL356967 Human DNA
9	21	1.5	2286	2	AC013913	AC013913 Drosophila
10	21	1.5	55954	2	AL136322	AL136322 Human DNA
11	21	1.5	110000	2	AC091528_1	Continuation (2 of
12	21	1.5	129504	9	AL445469	AL445469 Human DNA
13	21	1.5	157913	2	AC106679	AC106679 Rattus no
14	21	1.5	171926	2	AC084374	AC084374 Homo sapi
15	21	1.5	176912	2	AC096964	AC096964 Rattus no
16	21	1.5	212888	9	AL513325	AL513325 Human DNA
17	21	1.5	262059	3	AE003573	AE003573 Drosophila
18	21	1.5	303040	1	AP000991	AP000991 Thermopila
19	20	1.4	624	10	MMVH81XR2	U16744 Mus musculu
20	20	1.4	1208	8	AY010964	AY010964 Hordeum v
21	20	1.4	1218	8	AY010965	AY010965 Elymus el
22	20	1.4	1218	8	AY010966	AY010966 Elymus g1
23	20	1.4	1218	8	AY010967	AY010967 Elymus g1
24	20	1.4	1218	8	AY010968	AY010968 Elymus g1
25	20	1.4	1218	8	AY010971	AY010971 Elymus r1
26	20	1.4	1218	8	AY010975	AY010975 Elymus v1
27	20	1.4	1218	8	AY010976	AY010976 Elymus v1
28	20	1.4	1218	8	AY010977	AY010977 Elymus wa
29	20	1.4	1218	8	AY010978	AY010978 Elymus wa
30	20	1.4	1219	8	AY010961	AY010961 Hordeum b
31	20	1.4	1242	8	AY010981	AY010981 Elymus g1
32	20	1.4	1247	8	AF331953	AF331953 Elymus v1
33	20	1.4	3235	10	RATNOP140B	M94288 Rattus norv
34	20	1.4	3609	10	RATNOP140A	M94287 Rattus norv
35	20	1.4	10310	10	AC084885	AC084885 Cfancg/X
36	20	1.4	33758	9	CEM79	250806 Caenorhabdi
37	20	1.4	36823	9	HSU209G1	268873 Human DNA s
38	20	1.4	40142	2	AC015137	AC015137 Drosophila
39	20	1.4	49438	2	AC106585	AC106585 Rattus no
40	20	1.4	72207	9	AL596254	AL596254 Human DNA
41	20	1.4	74328	2	AC100878	AC100878 Mus muscu
42	20	1.4	102104	2	AC106430	AC106430 Rattus no
43	20	1.4	121067	2	AC096119	AC096119 Rattus no
44	20	1.4	135489	9	AC026446	AC026446 Homo sapi
45	20	1.4	139887	9	CNS01RGX	AL160233 Human chr

ALIGNMENTS

RESULT 1
AX119931
LOCUS AX119931 12793 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 1 from Patent WO0129266.
ACCESSION AX119931
VERSION AX119931.1 GI:14036678
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 12793)
AUTHORS Hudson,T.J., Engert,J. and Richter,A.
TITLE Identification of arsacs mutations and methods of use therefor
JOURNAL Patent: WO 0129266-A 1 26-APR-2001;
MCGILL UNIVERSITY (CA) ; Hopital Sainte-Justine (CA)

FEATURES
source
1..12793
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 4163 a 2256 c 2487 g 3887 t
ORIGIN

Query Match 92.4%; Score 1295; DB 6; Length 12793;
 Best Local Similarity 100.0%; Pred No. 0;
 Matches 1295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtagcagtaaaactagagcagtcaccaagcgacacaaagccttagaagaatagatgcc 60
 |||||||
 Db 5300 GTAGCAGTAAACTAGAGAGAGTCCCAAGCGACACAAAGCCTTAGAAGATATGCAATCC 5359
 QY 61 aatgctgttttcaaacacttggcacaagaatttgggcagaagaataatggaccagcaga 120
 |||||||
 Db 5360 AATGCTGTTTTACAACACTTGGCACAGAAATTTGGCGAGAAAGAAATTTGACCACAGAGA 5419
 QY 121 attaagagcactcttaataatgatactctctgaagaagaatttgaagaagctcttccaa 180
 |||||||
 Db 5420 ATTAAAGCATCTCTTAATGATATCTCTTGGAAAAGCAATTTGAAAGAGCTTCTTCA 5479
 QY 181 aatgctgatatgacaaaggcgacagaatctgtttgtgttgcatacctagacagcatcca 240
 |||||||
 Db 5480 AATGCTGATGATGCAAAAGGCGACAGAAATCTGTTTTGTGTTGATCCTAGACGATCCA 5539
 QY 241 gttagatagaatatttgatagaatgtggcccaattgcaaggccagcacttgtgtgtac 300
 |||||||
 Db 5540 GTTGAATAGAAATATTTATGATGATAGTGGGCCCATTTGCAMGGCCAGCACTTGTGTGTAC 5599
 QY 301 aacaaccagcactttacagagaatgagttagagaatcagaatcttgaagaaggcagcag 360
 |||||||
 Db 5600 AACACACAGCATTTTACAGAGATGATGTTAGAGAAATTCAGAAATCTTGGAAAAGCAGC 5659
 QY 361 aagaagggaatcctataaaactgagcagatagagaatagatccaattctgtgtacat 420
 |||||||
 Db 5660 AAGAGAGGAATCTTATATAAACTGAGACAGTATGAAATAGATCAATCTGTGATCAT 5719
 QY 421 atacagcagcagccacttttatttctggaatagacatcctgtgtatttttgatcctcat 480
 |||||||
 Db 5720 ATACAGAGAGCCCATCTTTATTTCTGGCAAGACATCCTGTGATTTTGTGATCCCAT 5779
 QY 481 gccagatatgacacaggggscacatccatagtcgccgagcagctgttttagagaattgcat 540
 |||||||
 Db 5780 GCCAGATATGACACAGGGGCCACATCATTAAGTCCCGAGAGCATGTTTAGAGATTTGAT 5839
 QY 541 gcagattttaagacacagttctcagabgtctgagatccttatcttgggaaccatlttaaa 600
 |||||||
 Db 5840 GCAGATTTTATGAGACACAGTTCTCAGATGTTCTGATCTTATCTGGGAACCCATTTTAA 5899
 QY 601 ctggaataatggcacaatggttcagatttctctctgtaatgacaaatgggaagaatttgcg 660
 |||||||
 Db 5900 CTGGATTAATGGCAAAAGTTCAGATTTCTCTTCTGTAATGCAGAAATGCGAAAAGTTTCG 5959
 QY 661 gaaatttcgtctgttcacagatcagacagaatgctcagagatcttggacaacatgacg 720
 |||||||
 Db 5960 GAATTTTCGTCTGTTCACAGATCAGACAGAAATGTCAGAAATCTTTTGGCAAACTCGCC 6019
 QY 721 tcaagatggggcagaactcttaattgttcttaacacatggaaaaaattctatttggaa 780
 |||||||
 Db 6020 TCAGATGGGGCAGAAACTCTTAATGTTCTTAATCACATGAAAAAATTTCTATTGGA 6079
 QY 781 atagataagagatctgagctcttaaatgtgtctgatttcagtaaggcaaaatcacagat 840
 |||||||
 Db 6080 ATAGATTAAGAGTACTGAGCTCTTAATGTCTTATTTCACTAAAGGCAAAATCACAGAT 6139
 QY 841 ggaagacagattgaagaagaacaattctatgactctgttaattgatatgttactaaag 900
 |||||||
 Db 6140 GAGACAGANTGAAAAGGAAACAATTTTCATGCTCTGTAATTAATAGTGTACTTAAAG 6199
 QY 901 aggcagctcaaaagacataccagttcaacaataactatactatgatacttgagagactc 960
 |||||||
 Db 6200 AGGCAGCTCAAAAGACATACAGTTCAACAAATACCTATATGATGATCTAGGAGACTCT 6259
 QY 961 gaagagaatcttactacgtgtgactaaattgttaatatagatcaggtcttcaagtatgagaaa 1020
 |||||||
 Db 6260 GAAGGAATCTTACTACGTGGCTAAATTTGTAATAGATCAGGCTTTTCAAGTATGAGAAA 6319

QY 1021 gtatctaaagtgtcatatcagctcacaagaaccacaagatatactcttccacgtggt 1080
 |||||||
 Db 6320 GTATCTRAAAGTGTCATATAGCTCACAGACAGATTTACTCTTTCCACGCTGT 6379
 QY 1081 ggaatagctgctgcatlactcacaactataaaaccccatagggccttctgtttttg 1140
 |||||||
 Db 6380 GGAGTAGCTCCTGTCATACTACAACTATPAAAAAACCCTATAGGCGCTCTGTTTTTGG 6439
 QY 1141 cctcttcttcttggaagatggcgctgcatlctcagtgtaagtgccactttgacatgattca 1200
 |||||||
 Db 6440 CTTCTTTCTTTTGAAGACTGGGCTGCCATTTTCATGTAAGGCTCATTTGCACTGGATTTCA 6499
 QY 1201 gccagaagaaactgtggcgctgtaataatggagtggtgtctcgaagtgcagtaataac 1260
 |||||||
 Db 6500 GCCAGAAAGAACTGTGCGCTGATGATTAAGAGTTGTTGTTGGAAGTGAATGATGCAATAC 6559
 QY 1261 agttaatgacagaatataagctcctgcatatgt 1295
 |||||||
 Db 6560 AGTTAATGACACATTAATAGCTCCGATATGT 6594

RESULT 2
 AF193556 12793 bp DNA linear PRI 07-FEB-2000
 LOCUS AF193556 Homo sapiens saccin (SACS) gene, complete cds.
 DEFINITION AF193556
 ACCESSION AF193556
 VERSION AF193556.1 GI:6907041
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 1 (bases 1 to 12793)
 Engert,J.C., Berube,P., Mercier,J., Dore,C., Lepage,P., Ge,B.,
 Bouchard,J.P., Machieu,J., Melancon,S.B., Schalling,M.,
 Lander,E.S., Morgan,K., Hudson,T.J. and Richter,A.
 ARSACS, a spastic ataxia common in northeastern Quebec, is caused
 by mutations in a new gene encoding an 11.5-kb ORF
 Nat. Genet. 24 (2), 120-125 (2000)

JOURNAL
 MEDLINE
 2 (bases 1 to 12793)
 Engert,J.C., Berube,P., Dore,C., Lepage,P., Ge,B., Hudson,T.J. and
 Richter,A.
 Direct Submission
 Submitted (08-OCT-1999) Genome Centre, Montreal General Hospital,
 1650 Cedar Ave., Montreal, QC H3G 1A4, Canada
 Location/Qualifiers

FEATURES

source
 1..12793
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="13"
 /map="between D13S232 and D13S292"
 1..12793
 /gene="SACS"
 /product="sacsin"
 1..12793
 /gene="SACS"
 77..11566
 /gene="SACS"
 /note="molecular chaperone"
 /codon_start=1
 /product="sacsin"
 /protein_id="AAF31262.1"
 /db_xref="GI:6907042"
 /translation="MNTFPPGRELIVQWYPPDENRNPDSVSLKAWKNLYTHSEDL
 TLFDEMPILPRLIEGQVCLIRLRPSLYLIDDESAPQLPEFLADIVOLGFEVL
 KRIDASIQHPLIKKIYHSPVAVQIMKMYLQKCNQITSLPFRKDALRKFASL
 TPSSSEKRIIOELAIIFKRINNSDQISSYTKLGCKVLLHTTALPADLRISVID
 SSDPAIRLANMLKTEOLKTSCTLYIKDINAFYSHEVYQOLMAMVENSISIKNE
 NPNVBLPLPLFKFIOISQEQMSAGLEPDPIDVLEKDFCNEEGTTFPPSVTSFDIL
 HSLRQIGKNEASLKEKDVQYAKRILEAQVACPDQVLLKRAKATLLVLKRNHTLL
 OSSEKMTLKKIKVYACKERPKNYSGSLVMWGDLCNLCAPDKMDVGHAILIGSLP
 LVESIHVNLEKALGIFTRPSLSAVLKHFKIIVDVWSSKTFSDSDYYQFHIILLEYGF

MDHLENGKDSFALRFPWYTGKKFPCPLQAAVTKPIHDLQPYLHNPVTMAKFFHO
LFKVGCSIELTSDHISMVIOKILYKSDSDLOSQESONILHMLNITRMJLSONIPAS
PNTVPPIHHSKNSKSLIMKPIHBCCYCDIVDDLDELDSBPVITLHVDHDMKPTAE
MLKVPCLSTRLINPENMGFESQOREPIYTRIKINILEEYVSVDITFELLQNMADANA
TECSFLIDMRNMDIRENLDPGMAACHGALSNFNSQFSDSPVNTLIGSLKKG
EDVKVGFDMGFSYVHTDIPIMSEKEFIMDPNINHISKIKDSQKIRIIMSKG
OORLRKFPNOEPFIDVFCQLPLTVEAAYNGTLEFSLFSPFOQAKSESTCY
MTADISLVEFSLGCHRLIFITOSVSMYKTLKIETNPSLAODIVIIKKSSCSK
ALNTPVLSYLKAAKMLKTCSSSNKRLPSPEKSSCIOTIYVEERFAPNLEIDTA
LEFGPPDDPAALFEAKSGSKRPDELSKPTVCTTWTLCMDTEALRFSISESG
RRLGLVPCGAVGVLSLIDQKMTVPKPHIEVECTYLPRLKRTGLPVIINCCFAVTSR
KEIMKIDGTGRNMTFMHRYIVKAYLQVLSVRLDLAAGSNLTYTAYVAPDDPLVD
DFEIVCOGEDYEDIAHMGKELTVEFSDGYSWMSKNRFLDSDILKRDYDAAFLF
LKTKTGSKNICLAVELSSVKGFEAGCKOILEMTEFSEKOFESFVPNPLEIDTA
ELRDPMLTPIYLNKVDSESVLRTPCISLGHPIRPLTSLHPGRKXKLEIDT
GRPPYSTODIYLPILIKVOLGMAKDDLMDMLERANSVAEINKSDVAAICRS
LILSLIDERLKTIDPRAKDPAAYQITRPLPLRPGSLDMKNGSEKFEPTMAAD
LYTAEHODIVCLQPLINENSHFRCGGSYLAVERFLGKRPYDVLVINOKEVAK
SYVDGITLYXENTNACYKYLHBMALMONEITKMSIDKLFPESLIVENAYOSEKS
FHLNFEAPLYLOLPNNKYNRELFETVETVROSCVTEDEALVESIDORGRKOITE
ENROLCRIITISEGMSLIREKROEFCGNKGLIDPPTNMLPAKSLCNDCPMIV
KOTVYKCHADIPREYAVKLGAVPKRHKALERTASNCFTTLGFEQKRLTSRIS
ILNAYPSEKEMLELLQNMADAKATEICFVDPQHVVDHIFDKMAPLOGPALCVN
NOPFTEDYVRGIONLKGTEGEPYKQYQIGFNSYHTDPSFISGDIICIEDP
HARYAGATSIISPRMRDLADPFRQFSEVLDLYCTHFKLMDCTMFRPRPLRAEMA
KVEISSVPASDRMONLIDKLRSDGELIMPLINHEKISICEIDKSTGALNLYSVK
GKITDDELKRRKOPHASVIDYVTKRQKRLPVOQITTYTMDTEDSEBNTWILCONS
GESMEKVSXSVISAHKNODITLPRGCVACITHNKPKHRAVECLPLLELGLPH
VNGHFLPDSARNLMDNDNGVAVSRDNNLSMLALPAVAVELLIDKKRYFPGSDP
LSVLOMTPIHVVADTLKFLSEFPVNLDDPDLCLYKALVNICIHEMRLPLPVVA
PNIDGSLHSAVITWTIMSTNKTREFPNHLODELQHKNDYNTTKRTVAENY
RLKHLLETFENLYVNDDETANLYHCLIDDIVSYVTPDIDISFMTPESSPBNCH
GKPCRLQOQNLKFLSHSLKLVYCRFDAENELVEBGDLITLDSVLOTFFDAKPK
PLTTTIELHPSKDLFMNLYLYKTSNLLCKVAKVDFISFADLSSVLPREKTKS
CTWKDNFASBSWLKNAMHPISESVKEDQEBTKPFDVLDLKMALPGKTFYV
SANOQVYBEGDVLPLSLMHAIVFPNOSDKVFNHALKACIOLALIKCSKSAFVP
LILSCHTANISEPTSLILKAHYNOTSFREKLEYENDELALMENCNLHMSODDI
KILKSLPYKYSIGRYVSIKFGTCYVLTISPSAEVEMKTOSSSAFLEREKIHKLK
XYEIGCVPNVDLEVYLKHLIPKTEINSYDAKLEHLIYLRKLSABEELSEKQEPK
LESLLITHDANSRLQKQKHFYDRTVRFVEVMBEKLIPDNDFERKQLILKPKNHVE
MTSWEFELRNIGKYLILSOOLQFAEIRSEVANTEMSKETLONTDVLHNFHFOR
MDLSGNLELSELIPFLPERAPAEIRPHPOYEJNGVGLPIKFGAUVNPFKOC
DYQLMLTSCPIPERKATPLSIKEQSGSDGPOBOLVONMLNLDPLDYLINCC
RNICNTITLDEEMVTKRAKVRKSIYEELISAERREPRQLGVAFVMEDEKILKPRE
VYINLEESDPKRYLYKLPLELSTPHOLFNHGTEDITISKQYEVUSIRFKNSEGO
LDNEMKTVARVVSGLFRSLQNSVAKRSDLENVRLATLPSQDGLVSSILVDE
APHYKSRIOGNIGVOMLVDSQYLGKDHGFHKLIMLPQKLRPLRLSILBEOIDE
EPPKVOFGALCSLQRLQILSSEOTITGLIRMKHENNAFLANEKARILCKALR
EGLKVSCEFEKLOTLTRKGFNPPIPHSSETEFAFKRGNVAILLYIOHNSKIDNELL
ALAMTLKATNDNLISDPSYILAMLGIDYIRIGEKIDSLCKVYKDSPPSKLEPMPT
PIPAEIHITLMDPMNVFYRGEYVGYLVDAEGGDYISOTPYTAYITLVOEDVEDD
NSSFLEKTIYDIDIGYSEYKIVSSLDLYKFSRPESSOSRSABSTPSPSEFLTPGLR
SIPPLFSGRSHKTSKHOQPKRLKVNLSPEILKEVSVYVOMAKLDESERKAKIRRL
LYLKWHPDKNPNENDIANEYFKHLQNEINREKQAFLOONDAISRRFISASFOG
KYSFORFYTSAMNOEATSHKGEROOONCEKPPSAGIYOSRFPYPPFKFVGNVLA
RMTROARANSFARNDLHKNAEMVCKCTSLKXIALADYAVRGKSDSDYPTAAR
OKTEBEYSQOLEGLTNDVHTIEATGVDSLKTRYDLDLPFOIPDNDRFSEVAMRYMECT
ACITIKLENMOQKV

BASE COUNT 4163 a 2256 c 2487 g 3887 t
ORIGIN

Query Match 92.4%: Score 1295; DB 9; Length 12793;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacagtaaacagagcagtcaccaagagcacacaagccttagaagaatagcatcc 60
DB 5300 CTACAGCTAAACCTAGACGAGCTCCCAAGCAGACAAAGCCTTAGAAGATATGCACTCC 5359
QY 61 aatgctgctttaccacaacttggcacagaatttgggcagaaagaaatgacacagaga 120
DB 5360 AATGCTGCTTTTACACACTTGCGCAGAAATTTGGCGAAGAAATGATGACACAGAGA 5419

QY 121 attaagacatcccttaataagcatatccctctgaaagaaatgctgaaagcctcttcaa 180
DB 5420 ATTAAGACATCTCTTAATGCAATATCTTCTGAAAAGAAAGTGTGAAGACGCTTTTCAA 5479
QY 181 aatgctgctgacgaaagcgacagaaatcctgcttctgcttctgacccagacatcaa 240
DB 5480 AATGCTGATGATGCAAGCGCAGAAATGCTTGTGTTGTGATTCCTAGACAGATCCA 5539
QY 241 gttgataaataattgataagaaatggtggcccatctgcaaggccagcacttgctgac 300
DB 5540 GTTGATGAATAATTTGATGATAAGTGGGCCCCCATTCGCAAGGCCGACCTTGTCTGAC 5599
QY 301 aacacccgcatttaacgaagaatgatagtttgaggaattcgaactctggaaaaggcgc 360
DB 5600 AACACCCGACCATTTACGAAGAATGATGTTAGAGGAATTCAGAACTTGGAAAAGCCACG 5659
QY 361 aaagaggaatacccttaataaacctgacagataggaatgaattcctgctgatacat 420
DB 5660 AAAGAGGAATCCTTATTAATAAAGTGCAGATGGAATGGAATTCATTTCTGTATCAT 5719
QY 421 atcacagactgcccactcttcttaattctgagcaatgacatccctgctgataattcctcat 480
DB 5720 ATCACAGACTGCCCATCTTTATTTCTGGCAATGACATCTGTGATTTTGTGATCTCAT 5779
QY 481 gccagataatgacccaaggcgccacatccatgctcccggaagcagatgttaagagattgat 540
DB 5780 GCCAATATATGACACCGGGGCCACATCCATTAATGTCGGAGCGATGTTAGAGATTTGGAT 5839
QY 541 gcaagatttaagacaacattctcagatgttcctgatactctatctggaaccattctaa 600
DB 5840 GCAGATTTTAGACACACTTCTCAGATGTTCTGATCTTTATCTGGAAACCATTTTAAA 5899
QY 601 ctgataatctgacaaatgttcagatcttccctctctgaaatgcaagaaatggcaaatctgc 660
DB 5900 CTGATATATTCACAAATGTCAGATTTCTCTCTGATGCAAGAAATGGCAAAATTTTCG 5959
QY 661 gaaattctgctgtctcagacatcagacagaatgctgcgaatctttgacaaactctgc 720
DB 5960 GAAATTTGCTGTGTCGACGATTCACAGAAATGTCGCAAGATCTTTTGACCAAACTGTGCG 6019
QY 721 tcaagtgggacgaactcttaatgtcttctaatacacaatgaaataactcttaattctgaa 780
DB 6020 TCAGATGGGGCAGAACTTCTAATGTTTCTTAATCAGATGGAATAATTTCTAATTTGTCAA 6079
QY 781 atagataagactgagactcttaaatgctgataatcaagaaaggcaaaatcacagat 840
DB 6080 ATAGATAAGACTGAGACTCTTAATGCTGTATTCAGTAAAGGCAAAATCACAGAT 6139
QY 841 ggaagcaagatgaaagaaacaattctatgcatctgtaattgataagtgcttaactaaaag 900
DB 6140 GGAGCAGATTTGAAAAGGAAACAATTTATCATGTCATGTAATTTGATGATGTTAAAG 6199
QY 901 aggcagcctcaagacataccagttcaacaataactactatgataactgagagactc 960
DB 6200 AGGCAGCTCAAGACATACGACTTCAACAATAATACCTTACTATGATACAGAGACTCT 6259
QY 961 gaaagaaatcttaactagctgagcttaattgtaataagatacaggctttcaagiatgagaa 1020
DB 6260 GAAGGAATCTTACTACGTGCTAATTTGTAATAGATCAGGCTTTTCAAGTATGAGAAA 6319
QY 1021 gtaataaagtatcatatcatgctcaagaagacaaagatatactctttcccaagtgct 1080
DB 6320 GTATCTTAAAGTGTATATATGACGCTTCAAGAACCAAGATATTACTCTTTTCCACAGTGTGT 6379
QY 1081 ggaatagctgctgctattacacacaactataaaacccataaggcctctctgctttctg 1140
DB 6380 GGAGTAGCTGCTGCTATTAATCAGACACTATAAAAAACCCATAGGCGCTTCTGTTTGTG 6439
QY 1141 cctcttcttggagactggcctgcatcttcatgtaagtgaatggcacttgcacatgata 1200
DB 6440 CCTCTTCTTGGAGACTGGCGCTGCATTTTCATGTGATGGCCACTTTCGCACTGATGATCA 6499

Qy 1201 gccagaagaccctgtgcgtatgataatgagctgtgtctcgaagtcagctgaaacac 1260
|||||
Db 6500 gccagaagaccctgtgcgtatgataatgagctgtgtctcgaagtcagctgaaacac 6559
Qy 1261 agcttaatgacagcattaatagctctcatatgt 1295
|||||
Db 6560 agcttaatgacagcattaatagctctcatatgt 6594
RESULT 3
AL157766/c 92693 bp DNA linear PRI 11-APR-2001
LOCUS
DEFINITION Human DNA sequence from clone Rpl1-40020 on chromosome
13q12.11-12.2, complete sequence.
ACCESSION AL157766
VERSION AL157766.9 GI:13620292
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Tsimans, A.
TITLE Direct Submission
JOURNAL Submitted (11-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT
On Apr 12, 2001 this sequence version replaced gi:12709868.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; SW:
SWISSPROT; Tr: TrEMBL; Wp: WormPEP; information on the WormPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
Rpl1-40020 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/Bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
Rpl1-40020. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone Rpl1-760M1 is at 92594 in this sequence.
The true right end of clone Rpl1-72P19 is at 100 in this sequence.
FEATURES
source
1. :92693
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="q12.11-12.2"
/clone_lib="RPCI-11.1"
/clone="Rpl1-40020"
2390..2485
/note="MIR repeat: matches 81..192 of consensus"
repeat_region
2562..2673
/note="MIR repeat: matches 68..183 of consensus"
repeat_region
3896..4201
/note="AluY repeat: matches 3..308 of consensus"

repeat_region
5122..5397
/note="MER46C repeat: matches 1..286 of consensus"
repeat_region
18986..19294
/note="AluY repeat: matches 1..310 of consensus"
repeat_region
19644..19873
/note="MER46A repeat: matches 1..235 of consensus"
repeat_region
20613..20912
/note="AluSg1 repeat: matches 1..300 of consensus"
repeat_region
23342..23651
/note="AluSg1 repeat: matches 1..309 of consensus"
repeat_region
24769..24891
/note="L2 repeat: matches 2554..2682 of consensus"
repeat_region
25871..26011
/note="L2 repeat: matches 2356..2495 of consensus"
repeat_region
26033..26109
/note="L2 repeat: matches 2601..2688 of consensus"
repeat_region
26245..26344
/note="L2 repeat: matches 2154..2255 of consensus"
repeat_region
26938..27096
/note="MIR repeat: matches 3..175 of consensus"
repeat_region
27150..27653
/note="L2 repeat: matches 1063..1644 of consensus"
repeat_region
28522..28891
/note="THE1B repeat: matches 1..364 of consensus"
repeat_region
29447..29834
/note="L1ME3A repeat: matches 5787..6164 of consensus"
repeat_region
36098..36415
/note="AluSx repeat: matches 1..308 of consensus"
repeat_region
37202..37414
/note="MIR repeat: matches 22..262 of consensus"
repeat_region
37963..38254
/note="AluSg repeat: matches 9..301 of consensus"
repeat_region
38703..39008
/note="AluSg repeat: matches 1..306 of consensus"
repeat_region
39790..40093
/note="AluSx repeat: matches 1..304 of consensus"
repeat_region
40126..40416
/note="AluSg repeat: matches 1..292 of consensus"
repeat_region
40444..40733
/note="AluSg repeat: matches 1..292 of consensus"
misc-feature
41322..41405
/note="Single clone region. Assembly confirmed by
restriction digest data."
repeat_region
41541..41788
/note="AluSg repeat: matches 1..248 of consensus"
repeat_region
44790..45101
/note="AluSg repeat: matches 1..313 of consensus"
repeat_region
45261..45312
/note="13 copies 4 mer tggc 88% conserved"
repeat_region
45899..46206
/note="AluY repeat: matches 1..307 of consensus"
repeat_region
46754..47052
/note="AluY repeat: matches 1..298 of consensus"
repeat_region
47067..47365
/note="AluY repeat: matches 1..299 of consensus"
repeat_region
47477..47873
/note="L1MA10 repeat: matches 5950..6322 of consensus"
repeat_region
47889..48229
/note="AluSx repeat: matches 1..312 of consensus"
repeat_region
49168..49212
/note="Alu repeat: matches 85..126 of consensus"
repeat_region
49620..49693
/note="L2 repeat: matches 1685..1757 of consensus"
repeat_region
50704..51032
/note="AluSc repeat: matches 1..308 of consensus"
misc-feature
52204..53009
/note="CpG island"
evidence-not_experimental
53978..54137
/note="AluSx repeat: matches 1..160 of consensus"
repeat_region
54179..54511
/note="L1MB6 repeat: matches 5822..6172 of consensus"
repeat_region
55685..55949

repeat_region /note="AluX repeat: matches 1. .305 of consensus"
57331. .57390 /note="30 copies 2 mer ga 75% conserved"
repeat_region 57357. .57392
/note="9 copies 4 mer gaga 91% conserved"
repeat_region 57753. .57930
/note="MIR repeat: matches 82. .262 of consensus"
repeat_region 58260. .58389
/note="MIR repeat: matches 2. .153 of consensus"
repeat_region 58564. .58611
/note="24 copies 2 mer ca 93% conserved"
repeat_region 59350. .59533
/note="AluX repeat: matches 129. .313 of consensus"
repeat_region 59922. .60223
/note="AluX repeat: matches 85. .299 of consensus"
repeat_region 61036. .61144
/note="L2 repeat: matches 2581. .2696 of consensus"
repeat_region 62008. .62187
/note="TIGER1 repeat: matches 2238. .2418 of consensus"
repeat_region 62188. .62316
/note="AluX repeat: matches 1. .129 of consensus"
repeat_region 62330. .62363
/note="Alu repeat: matches 261. .294 of consensus"
repeat_region 62362. .62565
/note="TIGER1 repeat: matches 1586. .1787 of consensus"
repeat_region 62566. .62865
/note="AluX repeat: matches 1. .302 of consensus"
repeat_region 62866. .64385
/note="TIGER1 repeat: matches 46. .1586 of consensus"
repeat_region 64386. .64694
/note="AluX repeat: matches 1. .306 of consensus"
repeat_region 64695. .64713
/note="TIGER1 repeat: matches 29. .46 of consensus"
repeat_region 65068. .65395
/note="LIP2 repeat: matches 5405. .5733 of consensus"
repeat_region 65396. .65569
/note="AluX repeat: matches 136. .309 of consensus"
repeat_region 65571. .65640
/note="LIP2 repeat: matches 5728. .5791 of consensus"
repeat_region 65696. .65717
/note="11 copies 2 mer ta 100% conserved"
repeat_region 65725. .66096
/note="LIP2 repeat: matches 5789. .6155 of consensus"
repeat_region 66371. .66410
/note="10 copies 4 mer tctg 82% conserved"
repeat_region 67586. .67886
/note="AluX repeat: matches 1. .299 of consensus"
repeat_region 69748. .69930
/note="MIR repeat: matches 6. .248 of consensus"
repeat_region 70957. .71267
/note="AluX repeat: matches 1. .311 of consensus"
repeat_region 71279. .71413
/note="MER2B repeat: matches 548. .680 of consensus"
repeat_region 71411. .71737
/note="MER3A repeat: matches 47. .485 of consensus"
repeat_region 71780. .72075
/note="AluX repeat: matches 1. .295 of consensus"
repeat_region 72145. .72256
/note="MER31-internal repeat: matches 42. .175 of consensus"
repeat_region 72454. .72865
/note="MER31-internal repeat: matches 332. .739 of consensus"
repeat_region 72873. .73249
/note="MER31-internal repeat: matches 883. .1261 of

Query Match 92.4%; Score 1295; DB 9; Length 92693;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgacagtaaaactagagcagtcacaaagcagacacaaagccttagaagatatgcatcc 60
DB 13610 GTAGCAGTAAACTAGAGCAGTCCTCCCAAGCAGACAAAGCCTTAGAAGATATGATCCTC 13551

QY 61 aatgctcttttacacacacttgcagacagaatttggcgaaagaaatctgaccagcaga 120
DB 13550 AATGCTCTTTTACACACACTTGGCAGACAAATTTGGCGAAGAAAGAAATTTACACAGCAGA 13491
QY 121 ataaagacatccttaataatgatalatccttctgaaagaaatgctgaaagacttctcaa 180
DB 13490 ATTAAGAGCATCCTTAATGATATCTTCTGAAAGAAAGTGTGAAGAGCTTCTTCAA 13431
QY 181 aatgctatgctgcaaaagcgacagaatctgcttctgcttctgcttctgcttctgcttct 240
DB 13430 AATGCTATGATGCAAAAGCGACAGAAATCTGTTTGTGATCTTACAGACATCTCA 13371
QY 241 gttgataaatatttgatgataagttggcccccttgcgaagggcgacacatttctgttac 300
DB 13370 GTTGATGAATATTGTGATGATGATGAGTGGCCCCCATTTGCAAGGCGCATTTGTGTGATC 13311
QY 301 aacacacagccatttacagaagatgatactgtaagaaattcagaatctgtaaaagcagc 360
DB 13310 AACACACAGCATTTACAGAAAGATGATGATGAGAAATTCGAAATTCGAAAGGCGACG 13251
QY 361 aaagaaggaaatccttaaaactgacagatgagaaatgaaatcgaattctgtatcat 420
DB 13250 AAAGAGGAAATCCTTATAAACTGACAGATGGAATAGGATTCAAATTCGTGTATCAT 13191
QY 421 atcacagatgcccattcttattcttctggaatgataatccgtgtattttgataccat 480
DB 13190 ATCACAGATGCCCATTCTTATTTTGTGGCAATGACATCCGTATTTTGTGATCTCAT 13131
QY 481 gccagataatgacacagggcgacacataatgctccgagcagcagctttagaatgtagat 540
DB 13130 GCCAGATATGACACAGGGCGACATCCATTTAGTCCGAGCCATGTTAGAGATTTGGAT 13071
QY 541 gcgaattttaggacacagcttctcagatgcttgcatacttctcgtggaaccatttaa 600
DB 13070 GCAATTTTGAAGACACAGTTCTCAGATGTTCTGATCTTTATCTGGGAAACCATTTTAA 13011
QY 601 ctgataatgacacaaatgcttgaatcttctccttctgtaagaaatggaagaaatgctt 660
DB 13010 CTGATATTTGACAAATGTTGATGATTTCTCTTCTGTAATGCAAAAGTTCG 12951
QY 661 gaaattcgtctgttccagacatcagacagaatggtccagaatcttctgacaaactgcgc 720
DB 12950 GAATTTCTGTGTCTCCAGCATCAGACAGATGCTCCAGATCTTTTGGCAAACTGGCC 12891
QY 721 tcagaatggggcagaactcttaatgcttcttaataatcagatggaaaaaattctatttggaa 780
DB 12890 TCAGATGGGCGAGAACTTCTTAATGTTCTTAATCAGATGCAAAATTTCTAATTTGTGAA 12831
QY 781 atagataagtagttagagctcttaaatgctgctgattctcagtaaaaggcaaatcagat 840
DB 12830 ATAGATTAAGATGCTGAGCTCTTAATGCTGTAATGCTGTAATGCTGTAATGCTGTAAT 12771
QY 841 ggaagacagatgaaagaaacaaatctcatcactgctgtaattgtaagtgtaactaaag 900
DB 12770 GGAGACAGATTGAAAGAAAGAAACATTTCTATCTGTAATGTAATGTAATGTAATGTAAT 12711
QY 901 aggcagctcaaaagacatacagcttcaacaaataactatactatgatactgatactgactct 960
DB 12710 AGGCAGCTCAAAAGACATACCGTTCAACAAATACTATATGATGATGATGATGATGATGAT 12651
QY 961 gaaagaaatcttactagctgcttaattgtaataatgatacagcttctcaagtaagagaa 1020
DB 12650 GAAAGAAATCTTACTAGCTGCTGCTAATTTGTATGATGATGATGATGATGATGATGATGAT 12591
QY 1021 gtactaaagtgcatatacagctcacaagaacacagatatctcttctccagctggt 1080
DB 12590 GTATCTAAAGTGTAT 12531
QY 1081 ggaatagctgctgcatatactacactataaaaaaacccatagggccttctgtttttg 1140
DB 12530 GGAATAGCTGCTGCTATCTACTACACACTATTAATAAAACCCCATAGGCGCTTCTGTTTTTGG 12471

Qy 1141 cctcttcttggagactgggtcgtccattcattgtaatgacacttgcacttgatca 1200
|||||
Db 12470 CCTCTTCTTGGAGACTGGGCTCCATTTCAATGTAATGCCACTTTCACATGGATTCA 12411
|||||
Qy 1201 gccagaagaacctgtggcgatgataatgagtggtgttcgaagtgcactggaataac 1260
|||||
Db 12410 GCCAGAGGAAGACTGTGGCTGTGATGAATGAGGTGCTTGGAAGTGAAGTGAATAAC 12351
|||||
Qy 1261 agttaatgacagcattatagctcctcctcatagt 1295
|||||
Db 12350 AGTTTAATGACACATTAACTCCTGCATATGT 12316
|||||

RESULT 4
AC079761/c 99819 bp DNA linear HMG-10-SEP-2000
LOCUS PROGRESS ***, 44 unordered pieces.
DEFINITION Homo sapiens chromosome UNK clone RP11-143G17, *** SEQUENCING IN
AC079761
AC079761.1 GI:10047966
VERSION HTG: HTGS_PHASE1.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 99819)
TITLE The sequence of Homo sapiens clone.
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 99819)
REFERENCE Waterston,R.H.
AUTHORS Direct Submission
TITLE Submitted (10-SEP-2000) Genome Sequencing Center, Washington
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information

* NOTE: This is a 'working draft' sequence. It currently
* consists of 44 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1137: contig of 1137 bp in length
* 1138 1237: gap of unknown length
* 1238 2538: contig of 1301 bp in length
* 2539 2638: gap of unknown length
* 2639 3976: contig of 1338 bp in length
* 3977 4076: gap of unknown length
* 4077 5355: contig of 1279 bp in length
* 5356 5455: gap of unknown length
* 5456 6757: contig of 1302 bp in length
* 6758 6857: gap of unknown length
* 6858 8570: contig of 1713 bp in length
* 8571 8670: gap of unknown length
* 8671 9925: contig of 1255 bp in length
* 9926 10025: gap of unknown length
* 10026 11426: contig of 1401 bp in length
* 11427 11526: gap of unknown length
* 11527 13266: contig of 1740 bp in length
* 13267 13366: gap of unknown length
* 13367 14794: contig of 1428 bp in length
* 14795 14894: gap of unknown length
* 14895 16054: contig of 1160 bp in length
* 16055 16154: gap of unknown length
* 16155 17395: contig of 1241 bp in length
* 17396 17495: gap of unknown length

* 17496 19287: contig of 1792 bp in length
* 19288 19387: gap of unknown length
* 19388 21294: contig of 1907 bp in length
* 21295 21395: gap of unknown length
* 21395 22944: contig of 1550 bp in length
* 22945 23045: gap of unknown length
* 23045 24421: contig of 1377 bp in length
* 24422 24521: gap of unknown length
* 24521 25870: contig of 1349 bp in length
* 25871 25970: gap of unknown length
* 25971 27230: contig of 1260 bp in length
* 27231 27331: gap of unknown length
* 27331 28779: contig of 1448 bp in length
* 28779 30893: contig of 2015 bp in length
* 30894 30993: gap of unknown length
* 30994 32460: contig of 1467 bp in length
* 32461 32560: gap of unknown length
* 32561 33984: contig of 1424 bp in length
* 33985 34084: gap of unknown length
* 34085 35285: contig of 1201 bp in length
* 35286 35385: gap of unknown length
* 35386 37184: contig of 1799 bp in length
* 37185 37285: gap of unknown length
* 37285 39172: contig of 1888 bp in length
* 39173 39272: gap of unknown length
* 39273 40874: contig of 1602 bp in length
* 40875 40974: gap of unknown length
* 40975 42893: contig of 1919 bp in length
* 42893 42993: gap of unknown length
* 42994 44384: contig of 1391 bp in length
* 44385 44484: gap of unknown length
* 44485 45999: contig of 1515 bp in length
* 46000 46099: gap of unknown length
* 46100 48659: contig of 2570 bp in length
* 48670 48769: gap of unknown length
* 48770 50798: contig of 2029 bp in length
* 50799 50898: gap of unknown length
* 50899 52809: contig of 1911 bp in length
* 52810 52909: gap of unknown length
* 52910 55127: contig of 2218 bp in length
* 55128 55227: gap of unknown length
* 55228 58087: contig of 2860 bp in length
* 58088 58187: gap of unknown length
* 58188 61004: contig of 2817 bp in length
* 61005 61104: gap of unknown length
* 61105 64185: contig of 3081 bp in length
* 64186 64285: gap of unknown length
* 64286 67105: contig of 2820 bp in length
* 67106 67205: gap of unknown length
* 67206 70837: contig of 3632 bp in length
* 70838 70937: gap of unknown length
* 70938 75837: contig of 4900 bp in length
* 75837 75937: gap of unknown length
* 75938 80452: contig of 4515 bp in length
* 80453 80552: gap of unknown length
* 80553 84661: contig of 4109 bp in length
* 84662 84761: gap of unknown length
* 84762 90542: contig of 5781 bp in length
* 90543 90642: gap of unknown length
* 90643 94348: contig of 3706 bp in length
* 94349 94449: gap of unknown length
* 94449 99819: contig of 5371 bp in length.

FEATURES
source
1. 99819
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="UNK"
/clone="RP11-143G17"
1. 1137
/note="assembly_name:Contig20"
1338. 2538
/note="assembly_name:Contig23"

misc_feature 2639..3976
/note="assembly_name:Contig30"
misc_feature 4077..5355
/note="assembly_name:Contig35"
misc_feature 5456..6757
/note="assembly_name:Contig39"
misc_feature 6858..8570
/note="assembly_name:Contig40"
misc_feature 8671..9925
/note="assembly_name:Contig41"
misc_feature 10026..11426
/note="assembly_name:Contig42"
misc_feature 11527..13266
/note="assembly_name:Contig44"
misc_feature 13367..14794
/note="assembly_name:Contig47"
misc_feature 14895..16054
/note="assembly_name:Contig48"
misc_feature 16155..17395
/note="assembly_name:Contig49"
misc_feature 17496..19287
/note="assembly_name:Contig50"
misc_feature 19388..21294
/note="assembly_name:Contig51"
misc_feature 21395..22944
/note="assembly_name:Contig52"
misc_feature 23045..24421
/note="assembly_name:Contig53"
misc_feature 24522..25870
/note="assembly_name:Contig54"
misc_feature 25971..27230
/note="assembly_name:Contig55"
misc_feature 27331..28778
/note="assembly_name:Contig57"
misc_feature 28879..30893
/note="assembly_name:Contig58"
misc_feature 30994..32460
/note="assembly_name:Contig59"
misc_feature 32561..33984
/note="assembly_name:Contig60"
misc_feature 34085..35285
/note="assembly_name:Contig61"
misc_feature 35386..37184
/note="assembly_name:Contig62"
misc_feature 37285..39172
/note="assembly_name:Contig63"
misc_feature 39273..40874
/note="assembly_name:Contig64"
misc_feature 40975..42893
/note="assembly_name:Contig65"
misc_feature 42994..44384
/note="assembly_name:Contig66"
misc_feature 44485..45999
/note="assembly_name:Contig67"
misc_feature 46100..48669
/note="assembly_name:Contig68"
misc_feature 48770..50798
/note="assembly_name:Contig69"
misc_feature 50899..52809
/note="assembly_name:Contig70"
misc_feature 52910..55127
/note="assembly_name:Contig71"
misc_feature 55228..58087
/note="assembly_name:Contig72"
misc_feature 58188..61004
/note="assembly_name:Contig73"
misc_feature 61105..64185
/note="assembly_name:Contig74"
misc_feature 64286..67105

Query Match 65.9%: Score 923; DB 2; Length 99819;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtagcagtaaaactcaggagcagtcctccaaagcgacacaaagccttagaagaatagatgcatcc 60
Db 88196 GTAGCAGTAAACTAGAGGAGCGAGTCCTCCAAAGCGACACAAAGCGCTTAGAAGATATGCAATCC 88137
Qy 61 aatgtctgttttaaacaccttggcaccagaatttggcgagaagaagaatttaccacagaga 120
Db 88136 AATGCTGTTTTACACACCTTGGCACAGAAATTTGGCGAAGAAATTTGCCACGACGAGA 88077
Qy 121 ataaagacatccttaatgcatatccctctgaaaaagaaatgttgaagaagccttctca 180
Db 88076 ATTAAGACATCTCTTAATGCAATTCCTCTGAAAAAGAAAGTTGTAAGAGCTTCTTCA 88017
Qy 181 aatgtcgtatgtcagaagcgacagaatctgtttgtttgtttgatttaccagaacatcca 240
Db 88016 AATGCTGTATGTGCAAAAGCGACAGAAATCTGTTTTGTTTGAATCCATGACAGCATCA 87957
Qy 241 gttgataagaatatttgaatgaatgagtcggcccatgtgcaaggcgacagcacttgtgtac 300
Db 87956 GTTGATGAATATTTGATGATTAAGTGGGCCCATTTGCAAGGCGCCAGCACTTGTGTATC 87897
Qy 301 aacacccagccattacacagaagatgatttgaaggaattcagaatcttgaagaagcagc 360
Db 87896 AACAAACGACCATTTACAGAAAGATGATGTTAGAGGAATTCAGAAATCTTGAAAGACGACG 87837
Qy 361 aaagagggaaatccctataaaactggacagatgtaagtaagtaacatctcgtatcat 420
Db 87836 AAAGAGGAAATCCTTATTAATACTGGACATGATGGAATGAGATCAATTCGTGTATCAT 87777
Qy 421 atcacagcttgcacatcttatttcttgcgaatgacatccggtgatttggatctcat 480
Db 87776 ATCMAGACTGCCCTCTTTATTTCTGCAATGACATCTGTTGATTTTGTATCTCAT 87717
Qy 481 gccagatatgcaccaggggcacatccatagtcctccgagcagcatgtttagaatgtgat 540
Db 87716 GCCAGATATGACACAGAGGCGCACATCCATAGTCCGACGCGATTTAGAGATTTGAT 87657
Qy 541 gcaagatttgaagacagttctcagaatgttctgatacttctcgggaaccatcttaa 600
Db 87656 GCAGATTTTAGACACACATCTTCAAGATGTTCTGATCTTTATCTGGGAACCATTTTAAA 87597
Qy 601 ctgataatgacagaatgttcagatcttctctctgtaatgcaagaatggcaaaagtttcg 660
Db 87596 CTGATTAATTCACACATTTTACATTTCTTCTCTTAATGCAAGAAATGGCAAAAGTTTCG 87537
Qy 661 gaaatttcgtctgttcacgacatcagacagaatggttcagaaatctttagacaactgcgc 720
Db 87536 GAAATTTGCTGCTCCAGCATCAGACAGAAATGTCAGAAATCTTTTGACAAACTGGCGC 87477
Qy 721 tcagatggggcagaactcttaatgttcttaacatcagtaagaataattcattgtgaa 780
Db 87476 TCAGATGGGGCAGAACTTCTAATGTTTCTTAATCAATGGAATAATTTCTAATTTGTAA 87417
Qy 781 atagataagatctacgtggccttaaatgttctgtatataagaaagggaataatcagaat 840
Db 87416 ATAGATAAGATCTACGTGGCTTAATGTGCTGATTAAGTAAGAAAGGCAAAATTCACAT 87357
Qy 841 ggaagacaatgtgaaaaagaaacaaatltcatgcatctgttaattgtaagtgtaactaaag 900
Db 87356 GGAGACATTTGAAAGAAAGCAATTTTCATGCATCTGTAATTAATGATGTTAAAG 87297
Qy 901 aggaagctcaagaacatcacagt 923
Db 87296 AGGACGCTCAAGACATCACAGT 87274

RESULT 5
AX119933 11492 bp DNA linear PAT 11-MAY-2001
LOCUS AX119933
DEFINITION Sequence 3 from Patent WO0129266.
ACCESSION AX119933
VERSION AX119933.1 GI:14036679
KEYWORDS

SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Hudson, T.J., Engert, J., and Richter, A.
AUTHORS Identification of arscs mutations and methods of use therefor
TITLE Patent: WO 0129266-A 3 26-APR-2001;
JOURNAL McGill University (CA); Hopital Sainte-Justine (CA)
FEATURES
source
1. 11492
location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
BASE COUNT 3599 a 2280 c 2387 g 3226 t
ORIGIN
Query Match 3.8%; Score 53; DB 6; Length 11492;
Best Local Similarity 100.0%; Pred. No. 8e-17;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 298 tacacacaccagcatttacagaagatgatgttagaggaattcaagaatcttg 350
|||||
Db 5520 TACACACACCGCATTACAGAAGATGATGTAGAGGAATCAGAACTTGG 5572
|||||
RESULT 6
AF193557 11493 bp DNA linear ROD 07-FEB-2000
LOCUS Mus musculus sacsin gene, complete cds.
DEFINITION AF193557
ACCESSION AF193557 GI:6907043
VERSION AF193557.1
KEYWORDS house mouse.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 11493)
Engert, J.C., Berube, P., Mercier, J., Dore, C., Lepage, P., Ge, B.,
Bouchard, J.P., Mathieu, J., Melancon, S.B., Schalling, M.,
Lander, E.S., Morgan, K., Hudson, T.J. and Richter, A.
ARSCS, a spastic ataxia common in northeastern Quebec, is caused
by mutations in a new gene encoding an 11.5-kb ORF
Nat. Genet. 24 (2), 120-125 (2000)
20120709
2 (bases 1 to 11493)
Engert, J.C., Berube, P., Dore, C., Lepage, P., Ge, B., Hudson, T.J. and
Richter, A.
Direct Submission
Submitted (08-OCT-1999) Genome Centre, Montreal General Hospital,
1650 Cedar Ave., Montreal, QC H3G 1A4, Canada
Location/Qualifiers
1. 11493
/organism="Mus musculus"
/db_xref="taxon:10090"
<1. >11493
/product="sacsin"
1. 11493
/note="molecular chaperone"
/codon_start=1
/product="sacsin"
/protein_id="AAF31263.1"
/db_xref="GI:6907044"
translation="MTFWPGRGLVQWYFSEDRKRPISLMLKWKMKLYTHFSEDL
TLDEMPILPTLLNDQRCVELIRLRISVYLIDETFAOLPEPLADIVOKRGIVL
KRIDSIOHPLVKYKIHSPISAIIOIMKRILOKICNOIASILPHKRAKFLAST
TPTSEKRIIOELTFKRIINSDGISYTKLKGCKVLDHTAKIPTDLRSVSD
SSDEATIRLANMKTEKTKTSCLKFVLDENAFYTOEVQOLMKMILENLSLENE
NSNVLDMLPLKFIHMSQGVAAADLPDIEVLDELFEYNEEACFPPTITSPDIL
HSRLROIGLNESSEKEDVVOYARKEALQVSSCONQDVLMMKAKTLLIVLNKNTLL
OSSEGMALKIKWPACKERPNDGSLVWNGDCNLCAPDMCDAAHAYVGSSTP
LYESVAVNLEQALSTFTKPTINAVLKHFETVVDWTSKTFSEBDYIOFOHILLEITGF
MHDLHSEGDSEFRALKFPWWTGKNCPDLAOLAVIKFTHDLQPYLVNPKTMAKFKH

LFKAGSIEELTSDHISWYI0KYLYKSDQELSEESKONLHMLNIMWLYSNQIPAS
PMPVPIYHSRNPESKLMKPIHECCYCDIKVDLNDLDESEVILLVEODIPMKTAE
WIKVPCILSTRINPENMGQREPLRLEKVDLNDLDESEVILLVEODIPMKTAE
TCSFEMIDMRNMNDIRENLLDPMGACBAPALMSFENSSKIDNIRLDESLKRG
EYDKVRCFGLGRNSVHTIDPIIMREPMIMDPMINISKIHRSNPGIKIMSK
OQRKRKFPNPKPIIDVFCQPLAVEAPVSTNGLEFLSFRTDQOEAIVSEVSTCY
NMADIYSLDSESLGHLRILFQSVNSWYLYKLTIEFNPLADOTIIRKVCPSK
ALNAPVLSYLEAKLAKMKTCSSNNKLPDVPKSCILQITVEEFHVRRLADLOSP
LFPGPDPAATLEFEMAKSGOSKSDLEQKVTQCTWYLCICMDGELKESLNSG
RLGICPGCAVCYLHEQOQKWTVPVGHGECVPLIKIKGLPHINGCAVYNSR
KEIMKTDITGRNMTTMRVYIKAYIQALSVRLDLAGELIDYTYAYWPPDLYHD
DPSVICKGFEDIAHGKGLTRFVSDGSMWAMKVRFLDSDILORQVGAARFIE
LYIKTKTGSKNLCQAVLPSSVKAAGEEAKQILLETSEKQFSEVFPNIOELA
ELRDPMLNFVLEKELDEFGIILRTVPCVCSLEGLVLPVSRILHEGVAALFEDK
GREPVSTODYINPILIKRIVOLGMAKDIOIMDGMLEARSVAIRKSHQAALCSS
IISLIDELKIKIDPRKADPAKQYIPLPLTTRKAGSLFKMGKNSFPEMPAATD
ITYAEQDVCYCLLOPLNENSHSFRGCSVSLAVKFEGLCKLPYDYVINOLQVAK
SVDDGITVQENTNACVYKLEHVALQNEAKATTEKLEKPCFILLVENVYSEKVS
PHLNEAPAYLQLPNRYKNRRELSESVGRQSFVEFDALVLEIDREKQOITE
ENFOLCRITISEGIMSLIREKROEFCENYKGLIDPMLILLPAKSLCYNDCPWIKV
KOSTVYCHADIPREYAVKLGAIIPKRAKALERYANICETALGTEGOKELSTRIS
ILNAVPEKEMIKELLONADDAKATETICVPPROHPVDIRIDDKAPLOGLACVYN
NOPFTEDDVRGIONLCKGTREGPKCTGHGIGCFNVIHTTQSPSTISNDLIGTDP
HARYAPGATVSPGRNFRDLADPRTQFSVDLILYGNHFKLDCNTWFERPLNAMA
OVSSEISSVPSDRMVOQLDLKLSRDAEELMLNHEKISTICEIDKATGLVNLVSVK
GKITDGDRLKROKOFASVDSVYKROLIDPVOQITTYMDTEDSESGNLTYLMS
GFSSMEKVSQVISAHKMODITLPEPRGVAACITNHYKKHRAFCPLSLEGLPEF
VNGHFALDSARINMRDNGVGRSDMNSMTALIPALYVELLITOLKRRPGRSPT
LSYLONTPIHYVAKDTLKRFLSPPVARLDQDLYCLYALYALSCHEDKRRLPVYRA
PNDGSDLSHAYITWINSTSKTRPFEDNLQDLOLHLKNAIVITTRKVALENVY
RLKHLLEIGFNLVYNGDETANLHCLVADADIPVSVTPADVRSFMTSPSPDNCHI
GKLPRLQOOTNLKLFHSLKLVQDCEFKDAEESFEVEGLPLITLDSYQIDGKRPK
PLTYHELIPSRKDLFMTNLYLKYSSVLLNCVAKAVFSDISFADLSVLPREYKTN
CAKMDNFASSEWIKMAHFISESVSVDQDEPRKAPVIVYDIKMDALLGTRFTV
STISQLYVPEGDVILPLSLMHIAVFPNAGSDKPHALMGCIQOLANKICSDSALVP
LISCHTANIDSPASILKAVHVOVQTSFTEFKLEFALLMTCNLSHMSODDI
KILKSLPCYKISGRYMSIAKGTQCVLRSJPSAVERMOTSSSAFLEHMLKEL
YEYICGVPDVDELVEYLHLPLKIENTSYAKLELILYLNRLASLESEPEIKOLEAF
LESILITIHANNRNLKOKAHEVRYVVEFPEKFIKKEFKLEQVYIKRNOAF
MWSWPEFLNIGLKYALSOOLQLOPAKELSVANENKSPLEQSLVDLHHIOER
MDLSLNFLEKELSLIFLCPERAPATYIFPHOYOQVNTLPLIRKNGQVAPKRCOC
DVQQLSNFICPLIPERATPLSIREQGSOLAQEOLEQVLMNLINQVADLVLNVC
RINQNTITLDEEMVNRKAVLRHSIYEFLEAKREFFQILTKQVAFVWDEGMKLEFE
VINQENTLDEADFPYIKLPLELGTGHQLEKNGTEDIISTKOVEVLEIRIFSSSEKQ
LDPNEMRTAVRVSGLEFSLONDSYVRSDLEARDLALYLPQSDOKLVKSSILEVD
APHYKRSIOGNIGVOMLVDSOCYIGKHGPHFKILMLFPKRLPRLISLIEODE
EPPKYQFALCSLOGRDOLLSDEPTIGILIRIKHENDNAPLANEBEATILCAYLR
EGKLVSCFEKLDITLTVKGFNPLPHSRSTFAPLKRFGNAVILITIQHSDSKLNPGL
ALAMTLKASATDNLISDTSYLIAMGNDYIRISEKLDISAGVYDSEPSKLDLPMGT
PIPAEIHVTLMDPMNVFPGEVYGLVVAEGDYSQYQPYTYAIIQOEVEREDAD
NPSFLKITYOIDIGYSEKIVSGLDKFSRDESSONDSAPPTYSPEFLRTEGLR
SIPLESFGSKSHSPSTKHSRPKRLVNLPELKLQVSVDOAMKLPSEKRIIIR
LYIKHMPDKNPNHDIANEFKHLNKLNOKFLQNDVNOARASRRPSTASRPOS
DKYSRORTYSNNOEATSHKSERODSKKCPSPSAQOTYSQFPEPTPKSGNPEVA
RKMRLQARANFSAARNDLHKNANENWCFQYSLTKALALADYAVRAGSKDQVKFPAL
AQKIEEYSOOLEGLTINDVHTLAEYGVDSLKTRPDLPLPPQIIPNDRFSEVAMRWEC
TACIIIKLENFIOQKV"

BASE COUNT 3599 a 2281 c 2387 g 3226 t
ORIGIN
Query Match 3.8%; Score 53; DB 10; Length 11493;
Best Local Similarity 100.0%; Pred. No. 8e-17;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 298 tacacacaccagcatttacagaagatgatgttagaggaattcaagaatcttg 350
|||||
Db 5521 TACACACACCGCATTACAGAAGATGATGTAGAGGAATCAGAACTTGG 5573
|||||
RESULT 7
AC022025/c
LOCUS AC022025 189760 bp DNA linear PRI 15-DEC-2001

DEFINITION Homo sapiens chromosome 10 clone RP11-489D8, complete sequence.
ACCESSION AC022025
VERSION AC022025.6 GI:17861011
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 189760)
AUTHORS Smith,D.R.
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome
JOURNAL Sequence Data
REFERENCE 2 (bases 1 to 189760)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
REFERENCE 3 (bases 1 to 189760)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (15-DEC-2001) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
COMMENT On Dec 15, 2001 this sequence version replaced gi:12957675.
FEATURES
location/Qualifiers
1..189760
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-489D8"
/clone_1lb="RPCI-11"
BASE COUNT 61011 a 36277 c 33262 g 59210 t
ORIGIN
Query Match 1.6%; Score 23; DB 9; Length 189760;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1255 aataacagtttaatgacagcatt 1277
|||||
Db 184482 AATAACAGTTTAAATGACGACATT 184460
RESULT 8
LOCUS AL356967 163869 bp DNA linear PRI 09-FEB-2001
DEFINITION Human DNA sequence from clone RP11-427E4 on chromosome 6, complete
sequence.
ACCESSION AL356967
VERSION AL356967
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 163869)
AUTHORS Tracey,A.
TITLE Direct Submission
JOURNAL Submitted (08-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT On Feb 9, 2001 this sequence version replaced gi:11691497.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RP11-427E4 is from the library RPCI-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/dacpac/home.htm
VECTOR: pBACE3.6
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. This sequence is the
entire insert of clone RP11-427E4. The true right end of clone
RP11-685G11 is at 54518 in this sequence.

FEATURES

source
1..163869
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-427E4"
/clone_1lb="RPCI-11.2"
52..95
/note="5S repeat: matches 19. .65 of consensus"
203..250
/note="L2 repeat: matches 2463. .2510 of consensus"
340..461
/note="MIR repeat: matches 65. .192 of consensus"
467..953
/note="MT2C8 repeat: matches 1. .501 of consensus"
2103..2466
/note="L2 repeat: matches 2127. .2487 of consensus"
2658..2734
/note="L2 repeat: matches 1772. .1842 of consensus"
3382..3621
/note="L1PA4 repeat: matches 5907. .6146 of consensus"
3622..3815
/note="97 copies 2 mer aa 60% conserved"
7089..7132
/note="22 copies 2 mer ta 100% conserved"
7676..7778
/note="MIR repeat: matches 34. .138 of consensus"
8029..8907
/note="L1MA8 repeat: matches 288. .1284 of consensus"
8909..8974
/note="L1MA8 repeat: matches 4697. .4760 of consensus"
8975..9276
/note="Alusx repeat: matches 1. .302 of consensus"
9277..10794
/note="L1MA8 repeat: matches 4760. .6264 of consensus"
10806..11071
/note="L1MD repeat: matches 1278. .1535 of consensus"
11137..11260
/note="L2 repeat: matches 2579. .2710 of consensus"
12152..12571
/note="L2 repeat: matches 72. .517 of consensus"
13044..13865
/note="MER21B repeat: matches 1. .789 of consensus"
15518..15664
/note="MIR repeat: matches 57. .213 of consensus"
15761..16153
/note="MIR repeat: matches 1. .401 of consensus"
16284..16415
/note="Alu/Alu/PRAM repeat: matches 176. .312 of consensus"
17048..17352
/note="Alu repeat: matches 1. .305 of consensus"
17732..18043
/note="Alu repeat: matches 1. .298 of consensus"
18174..18393
/note="L1PA7 repeat: matches 5924. .6145 of consensus"

```

repeat_region 19046..19188
/note="MLT1J repeat: matches 129. .273 of consensus"
repeat_region 22980..23144
/note="L2 repeat: matches 2570. .2707 of consensus"
repeat_region 25307..25360
/note="27 copies 2 mer tt 72% conserved"
repeat_region 25620..25714
/note="L1M4 repeat: matches 6183. .6250 of consensus"
repeat_region 25715..25996
/note="AluJo repeat: matches 1. .282 of consensus"
repeat_region 25997..26027
/note="L1M8 repeat: matches 6250. .6285 of consensus"
repeat_region 26066..26259
/note="L1MC5 repeat: matches 7336. .7549 of consensus"
repeat_region 26522..26629
/note="MIR repeat: matches 76. .189 of consensus"
repeat_region 26767..27106
/note="MST1 repeat: matches 1. .394 of consensus"
repeat_region 27238..27847
/note="L1PA4 repeat: matches 5210. .5819 of consensus"
repeat_region 27848..28190
/note="L1PA4 repeat: matches 5800. .6142 of consensus"
repeat_region 29252..29372
/note="MIR repeat: matches 76. .200 of consensus"
repeat_region 30389..30552
/note="MER63A repeat: matches 3. .174 of consensus"
repeat_region 31093..31222
/note="MIR repeat: matches 13. .146 of consensus"
repeat_region 31943..31984
/note="L4 copies 3 mer tta 78% conserved"
repeat_region 32356..32545
/note="LTR33 repeat: matches 329. .521 of consensus"
repeat_region 32696..32897
/note="MER33 repeat: matches -10. .65 of consensus"
repeat_region 32994..32994
/note="LTR33 repeat: matches 90. .141 of consensus"
repeat_region 33597..33687
/note="MIR repeat: matches 56. .148 of consensus"
repeat_region 33726..33807
/note="L2 repeat: matches 2624. .2708 of consensus"
repeat_region 33843..33923
/note="27 copies 3 mer ata 65% conserved"
repeat_region 34008..34082
/note="MIR repeat: matches 66. .141 of consensus"
repeat_region 35184..35463
/note="AluJo repeat: matches 1. .288 of consensus"
repeat_region 36308..36729
/note="L2 repeat: matches 2229. .2695 of consensus"
repeat_region 37243..37308
/note="33 copies 2 mer aa 66% conserved"
repeat_region 38342..38636
/note="L1M4 repeat: matches 3889. .4188 of consensus"
repeat_region 38640..38678
/note="LTR40b repeat: matches 422. .462 of consensus"
repeat_region 38679..39027
/note="L1PA7 repeat: matches 5787. .6137 of consensus"
repeat_region 39028..39154
/note="LTR40b repeat: matches 308. .422 of consensus"
repeat_region 39159..39571
/note="L1M4 repeat: matches 2823. .3283 of consensus"
repeat_region 39626..39665
/note="L1M4 repeat: matches 2823. .3283 of consensus"
repeat_region 40095..40393
/note="20 copies 2 mer at 77% conserved"
repeat_region 40934..41565
/note="AluSc repeat: matches 5. .301 of consensus"
repeat_region 41925..42147
/note="L1MC repeat: matches 1203. .1921 of consensus"
repeat_region 42969..43268
/note="L2 repeat: matches 2466. .2745 of consensus"
repeat_region 43667..43708
/note="AluJb repeat: matches 1. .293 of consensus"
repeat_region 44902..45177
/note="21 copies 2 mer tg 100% conserved"

```

```

repeat_region /note="AluJo repeat: matches 1. .298 of consensus"
46349..46454
/note="53 copies 2 mer at 62% conserved"
repeat_region 47391..47625
/note="MIR repeat: matches 13. .262 of consensus"
repeat_region 47615..47663
/note="MIR repeat: matches 210. .252 of consensus"
repeat_region 47637..47678
/note="L2 repeat: matches 2708. .2750 of consensus"
repeat_region 50360..50596
/note="MIR repeat: matches 8. .261 of consensus"
repeat_region 51156..51205
/note="25 copies 2 mer ta 72% conserved"
repeat_region 51347..51733
/note="MLT1D repeat: matches 69. .500 of consensus"
repeat_region 51734..52036
/note="AluJb repeat: matches 1. .286 of consensus"
repeat_region 52037..52110
/note="MLT1D repeat: matches 2. .69 of consensus"
repeat_region 52237..52752
/note="L2 repeat: matches 1600. .2152 of consensus"
repeat_region 52768..53179
/note="LRR16A repeat: matches 25. .442 of consensus"
repeat_region 53191..53668
/note="L2 repeat: matches 2151. .2750 of consensus"
repeat_region 53672..53886
/note="MIR repeat: matches 19. .243 of consensus"

Query Match 1.6% Score 22; DB 9; Length 163869;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1129 ttctgttttgcctcttctt 1150
|||||
Db 94539 TTCTGTTTTCCTCTTCTT 94560

RESULT 9
AC013913 2296 bp DNA linear HTG 16-NOV-1999
LOCUS AC013913/C
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC013913
VERSION AC013913.1 GI:6437422
KEYWORDS HTG; HTGS. PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 2296)
Adams M. and Venter, J.C.
Direct Submission
Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
JOURNAL
COMMENT This sequence was identified as CDW:10213920 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1..2296
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 891 a 384 c 381 g 640 t
ORIGIN
Query Match 1.5% Score 21; DB 2; Length 2296;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1321 ggtattccctggtctgac 1341
 ||||||||||||||||||
 DB 1248 GGTATTCCCTGCTTGATC 1228

RESULT 10
 AL136322 55954 bp DNA linear PRI 04-JAN-2002
 LOCUS Human DNA sequence from clone RP11-75C23 on chromosome 1q31.2-32.1,
 DEFINITION complete sequence.
 ACCESSION AL136322
 VERSION AL136322.24 GI:18072465
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 55954)
 AUTHORS Smith,M.
 TITLE Direct Submission
 JOURNAL Submitted (04-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
 On Jan 6, 2002 this sequence version replaced g1:17902846.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em: EMBL; Sw:
 SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
 database can be found at
 http://www.sanger.ac.uk/Projects/C.elegans/wormep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
 Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr1
 RP11-75C23 is from the library RPCI-11.1 constructed by the group
 of Pieter de Jong. For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone
 RP11-75C23. It may be shorter because we sequence overlapping
 sections only once, except for a short overlap.
 The true left end of clone RP11-53124 is at 53955 in this sequence.
 The true right end of clone RP11-33218 is at 2000 in this sequence.
 FEATURES
 source
 1..55954
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="q31.2-32.1"
 /clone="RP11-75C23"
 /clone_1lb="RPCI-11.1"
 /clone_1lb="RPCI-11.1"
 BASE COUNT 17559 a 10247 c 10324 g 17824 t
 ORIGIN

Query Match 1.5%; Score 21; DB 9; Length 55954;
 Best Local Similarity 100.0%; Pred. No. 7.7;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1332 ggttcgacacacatatca 1352
 ||||||||||||||||||

DB 892 GGTTCGATCCACATTATCA 912

RESULT 11
 AC091528.1/c
 WPCOMMENT

Sequence split into 5 fragments LOCUS AC091528 Accession AC091528
 Fragment Name Begin End
 AC091528.0 1 11000
 AC091528.1 10001 21000
 AC091528.2 20001 31000
 AC091528.3 30001 41000
 AC091528.4 40001 431537
 Continuation (2 of 5) of AC091528 from base 100001 (AC091528 Homo sapiens chromosome

Query Match 1.5%; Score 21; DB 2; Length 110000;
 Best Local Similarity 100.0%; Pred. No. 6.9;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 227 ctgacagcatccagtgata 247
 ||||||||||||||||||
 DB 21772 CTGACAGCATCCAGTTGATA 21752

RESULT 12
 AL445469/c
 LOCUS Human DNA sequence from clone RP11-254N18 on chromosome 1, complete
 DEFINITION sequence.
 ACCESSION AL445469
 VERSION AL445469.4 GI:11991422
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 129504)
 AUTHORS Baaguley,C.
 TITLE Direct Submission
 JOURNAL Submitted (23-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
 requests: clonerequests@sanger.ac.uk
 On Dec 24, 2000 this sequence version replaced g1:11879926.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.

This sequence has been finished according to sequence map criteria
 as follows. An attempt is made to resolve all sequencing problems,
 such as compressions and repeats, but not necessarily within known
 annotated repeat sequence elements. Where the sequence is
 ambiguous, there is an annotation using the 'unseq' feature key.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C.elegans/wormep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
 Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr1
 RP11-254N18 is from the library RPCI-11.1 constructed by the group
 of Pieter de Jong. For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone
 RP11-254N18. It may be shorter because we sequence overlapping
 sections only once, except for a 100 base overlap.
 The true right end of clone RP11-254N18 is at 129504 in this
 sequence. The true right end of clone RP5-1180C10 is at 100 in this

sequence. The true right end of clone RP11-118D10 is at 95065 in this sequence.

FEATURES
Location/Qualifiers
1.129504
source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-254N18"
/clone_lib="RPCT-11.1"
42809.42924

misc_feature
/note="Sequence from overlapping clone ba118D10 (AL359938). Assembly confirmed by restriction digest."
BASE COUNT 44541 a 24211 c 23194 g 37558 t
ORIGIN

Query Match 1.5%; Score 21; DB 9; Length 129504;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1177 aatgcacattgcactgcat 1197
|||||
Db 107258 AATGCCACTTTCACCTGCAAT 107238

RESULT 13
AC106679/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-106A6,*** SEQUENCING IN PROGRESS
ACCESSION AC106679
VERSION AC106679.1 GI:18139203
KEYWORDS HTG: HTGS PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 157913)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alshrooks,S.L., Amartunge,H.C., Are,J.R., Banks,T., Barberia,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Cartron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Dem,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunartine,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsl,F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,U., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisedg,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Meli,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabat,K., Morgan,M., Morris,S., Moser,W., Neal,D., Newton,S.J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenkwo,S., Oguni,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojupokan,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shooshari,N., Sisson,I., Sodergren,E., Sonalle,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,

Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S., Williams,G., Williamson,A., Wleciyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.
Unpublished
2 (bases 1 to 157913)
Morley K.C.
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center Project name: GLUO
Center Clone name: CH230-106A6
Summary Statistics
findphrapList
Consensus quality: 143594 bases at least Q40
Consensus quality: 150694 bases at least Q30
Consensus quality: 156627 bases at least Q20
Estimated insert size: 142942; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-tp estimation
Quality coverage: 2.8x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 57 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1
5729: contig of 5729 bp in length
5730 5829: gap of unknown length
5830 12058: contig of 6222 bp in length
12059 12158: gap of unknown length
12159 12387: contig of 5219 bp in length
12387 17378: gap of unknown length
17378 17477: gap of unknown length
17477 23186: contig of 5709 bp in length
23186 23286: gap of unknown length
23286 28405: contig of 5119 bp in length
28405 33071: gap of unknown length
33071 33171: contig of 4566 bp in length
33171 33172: gap of unknown length
33172 33172: gap of unknown length
33172 37963: contig of 4791 bp in length
37963 38062: gap of unknown length
38062 42337: contig of 4274 bp in length
42337 42436: gap of unknown length
42436 42437: gap of unknown length
42437 47541: contig of 5105 bp in length
47541 47542: gap of unknown length
47542 47642: gap of unknown length
47642 51053: contig of 3412 bp in length
51053 51153: gap of unknown length
51153 54397: contig of 3244 bp in length
54397 54497: gap of unknown length
54497 57140: gap of unknown length
57140 57240: contig of 2643 bp in length
57240 57241: gap of unknown length
57241 60305: contig of 3065 bp in length
60305 60405: gap of unknown length
60405 64988: contig of 4583 bp in length
64988 65089: gap of unknown length
65089 65089: gap of unknown length
65089 69178: contig of 4090 bp in length
69178 69278: gap of unknown length
69278 71834: contig of 2556 bp in length
71834 71934: gap of unknown length

```

* 71935 74740: contig of 2806 bp in length
* 74741 74840: gap of unknown length
* 74841 77660: contig of 2820 bp in length
* 77661 77760: gap of unknown length
* 82112 82211: contig of 4351 bp in length
* 82212 82211: gap of unknown length
* 86065 86065: contig of 3854 bp in length
* 86066 86165: gap of unknown length
* 86166 88690: contig of 2525 bp in length
* 88691 88790: gap of unknown length
* 88791 91103: contig of 2313 bp in length
* 91104 91203: gap of unknown length
* 91204 94206: contig of 3003 bp in length
* 94207 94306: gap of unknown length
* 94307 95998: contig of 1692 bp in length
* 95999 96098: gap of unknown length
* 96099 99075: contig of 2977 bp in length
* 99076 99175: gap of unknown length
* 99176 102430: contig of 3255 bp in length
* 102431 104203: contig of 1673 bp in length
* 104204 104303: gap of unknown length
* 104304 106512: contig of 2209 bp in length
* 106513 106612: gap of unknown length
* 106613 109297: contig of 2685 bp in length
* 109298 109397: gap of unknown length
* 109398 113020: contig of 3623 bp in length
* 113021 113120: gap of unknown length
* 113121 115477: contig of 2357 bp in length
* 115478 115577: gap of unknown length
* 115578 117671: contig of 2094 bp in length
* 117672 117771: gap of unknown length
* 117772 119975: contig of 2204 bp in length
* 119976 120075: gap of unknown length
* 120076 121881: contig of 1806 bp in length
* 121882 121981: gap of unknown length
* 121982 124296: contig of 2315 bp in length
* 124297 124396: gap of unknown length
* 124397 125748: contig of 1352 bp in length
* 125749 125848: gap of unknown length
* 125849 127673: contig of 1825 bp in length
* 127674 127773: gap of unknown length
* 127774 129419: contig of 1646 bp in length
* 129420 129519: gap of unknown length
* 129520 130855: contig of 1336 bp in length
* 130856 130955: gap of unknown length
* 130956 132588: contig of 1633 bp in length
* 132589 132688: gap of unknown length
* 132689 134332: contig of 1644 bp in length
* 134333 134432: gap of unknown length
* 134433 136267: contig of 1835 bp in length
* 136268 136367: gap of unknown length
* 136368 137731: contig of 1364 bp in length
* 137732 137831: gap of unknown length
* 137832 139895: contig of 2064 bp in length
* 139896 139995: gap of unknown length
* 139996 141525: contig of 1530 bp in length
* 141526 141625: gap of unknown length
* 141626 143293: contig of 1668 bp in length
* 143294 143393: gap of unknown length
* 143394 144551: contig of 1158 bp in length
* 144552 144651: gap of unknown length
* 144652 146085: contig of 1434 bp in length
* 146086 146185: gap of unknown length
* 146186 147247: contig of 1062 bp in length
* 147248 147347: gap of unknown length
* 147348 148580: contig of 1233 bp in length
* 148581 148680: gap of unknown length
* 148681 149786: contig of 1106 bp in length
* 149787 149886: gap of unknown length
* 149887 150935: contig of 1049 bp in length
* 150936 151035: gap of unknown length
* 151036 152387: contig of 1352 bp in length

```

```

* 152388 152487: gap of unknown length
* 152488 153785: contig of 1298 bp in length
* 153786 153885: gap of unknown length
* 153886 155369: contig of 1484 bp in length
* 155370 155469: gap of unknown length
* 155470 156335: contig of 1166 bp in length
* 156336 156735: gap of unknown length
* 156736 157913: contig of 1178 bp in length.

Query Match      1.5%: Score 21; DB 2; Length 157913;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1018 aaagatctaaagtgcata 1038
Db 80552 AAAGATCTAAAGTGCATA 80532

RESULT 14
AC084374/c
LOCUS
DEFINITION
Homo sapiens chromosome 12q clone RP11-13403, WORKING DRAFT
ACCESSION
AC084374
VERSION
AC084374.20 GI:17105262
KEYWORDS
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 171926)
Muzny,D.M., Adams,C., Adlo-Oduola,B., Ali-Isman,F.R., Allen,C.,
Alshrocks,S.L., Amaralunga,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouk,T.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Cartron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,
Hollins,B., Homai,F., Howard,S., Huber,J., Huliyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., Kling,L., Korah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louisged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabadi,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkwo,S.,
Ogulu,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Raj,L.,
Oulles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Poile,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshbari,N.,
Sisson,I., Sodergren,E., Sonake,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Swalek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Tellford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Verra,Y., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,A., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wlasczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gdbbs.R.
Direct Submission
Unpublished
2 (bases 1 to 171926)

```

TITLE
JOURNAL
REFERENCE
2 (bases 1 to 171926)

AUTHORS
TITLE
JOURNAL

Worley, K.C.
Direct Submission
Submitted (27-OCT-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 27, 2001 this sequence version replaced g1:16874816.

COMMENT

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: HCCJ2
Center clone name: RP11-13403

Summary Statistics

Sequencing vector: Plasmid: M77789
Chemistry: Dye-terminator Big Dye: 99% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 173766 bases at least Q40
Consensus quality: 174661 bases at least Q40
Consensus quality: 175002 bases at least Q20
Estimated insert size: 172132; sum-of-coverage estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 6.4x in Q20 bases; sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source

1. 171926: contig of 171926 bp in length.
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12q"
/clone="RP11-13403"
/clone="31150 c 30825 g 52933 t

BASE COUNT
ORIGIN

Query Match 1.5%; Score 21; DB 2; Length 171926;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 ctgacgagcatcagttgata 247
|||||
Db 18991 CTGACAGCATCCAGTTGATA 18971

RESULT 15
AC096964
LOCUS

AC096964 176912 bp DNA linear HTG 20-DEC-2001
Rattus norvegicus clone CH230-202E24, *** SEQUENCING IN PROGRESS
DEFINITION
***, 67 unordered pieces.

AC096964 176912 bp DNA linear HTG 20-DEC-2001
AC096964.3 GI:17973319
HTG: HTGS_PHASE1.
NORWAY RAT.
Rattus norvegicus

SOURCE

Rattus norvegicus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS

1 (bases 1 to 176912)
Muzny, D.M., Adams, C., Adio-oduola, B., Ali-osman, F.R., Allen, C.,
Alsbrooks, S.L., Amarantunge, H.C., Are, J.R., Banks, T., Barbarella, J.,
Benton, J., Bimaye, K., Blankenburg, K., Bonnin, D., Bouck, J.,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buha, C.,
Burck, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
Ehaja, C., Escotto, M., Falls, T., Ferraruto, D., Flagg, N., Ford, J.,
Foster, P., Frantz, P., Gabisi, A., Geo, J., Garcia, A., Garner, T.,
Garza, N., Gill, R., Gorrell, J.H., Guevara, M., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
Hernandez, O., Hernandez, O., Hodson, A., Hognes, M., Holloway, C.,
Hollins, B., Homsi, F., Howard, S., Huber, J., Hulik, S., Hume, J.,
Jackson, L.F., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Kovach, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Lewis, L., Li, Z., Lichte, O., Lieu, C., Liu, J., Liu, W.,
Lounsbury, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M.,
Mei, G., Metker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokwenkwo, S.,
Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, R., Pickens, R., Primus, E., Pu, L.L.,
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M.,
Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshari, N.,
Sisson, I., Sodergren, E., Sontheimer, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Watlington, S., Williams, G., Williamson, A., Wleciak, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 176912)

Worley, K.C.

Direct Submission
Submitted (04-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced g1:17064280.

Center: Baylor College of Medicine

Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GHRN
Center clone name: CH230-202E24

Summary Statistics

Assembly program: Phrap: version 0.990329first call to
findphraplist

Consensus quality: 151663 bases at least Q40
Consensus quality: 160070 bases at least Q40
Consensus quality: 166999 bases at least Q20
Estimated insert size: 152238; sum-of-coverage estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 2.3x in Q20 bases; sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 67 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 7294: contig of 7294 bp in length
7295 7394: gap of unknown length
7395 12823: contig of 5429 bp in length
12824 12923: gap of unknown length
12924 17505: contig of 4582 bp in length
17505 17605: gap of unknown length
17605 24532: contig of 6927 bp in length
24532 24533: gap of unknown length
24533 29839: contig of 5207 bp in length
29839 29939: gap of unknown length
29939 35829: contig of 5890 bp in length
35829 35930: gap of unknown length
35930 39896: contig of 3967 bp in length
39896 44844: gap of unknown length
44844 44944: contig of 4848 bp in length
44944 50921: gap of unknown length
50921 51021: contig of 5977 bp in length
51021 54518: contig of 3497 bp in length
54518 54618: gap of unknown length
54618 59782: contig of 5164 bp in length
59782 59882: gap of unknown length
59882 63189: contig of 3307 bp in length
63189 63289: gap of unknown length
63289 67064: contig of 3775 bp in length
67064 70204: gap of unknown length
70204 70304: contig of 3040 bp in length
70304 73659: contig of 3355 bp in length
73659 73760: gap of unknown length
73760 76601: contig of 2841 bp in length
76601 81449: gap of unknown length
81449 81549: contig of 4749 bp in length
81549 84603: gap of unknown length
84603 84703: contig of 3054 bp in length
84703 87038: gap of unknown length
87038 87138: contig of 2335 bp in length
87138 88733: gap of unknown length
88733 88833: contig of 1595 bp in length
88833 92171: gap of unknown length
92171 92270: contig of 3337 bp in length
92270 94942: gap of unknown length
94942 95042: contig of 2672 bp in length
95042 97980: gap of unknown length
97980 98080: contig of 2938 bp in length
98080 101026: gap of unknown length
101026 101126: contig of 2946 bp in length
101126 103481: gap of unknown length
103481 103581: contig of 2355 bp in length
103581 105942: gap of unknown length
105942 106042: contig of 2361 bp in length
106042 108765: gap of unknown length
108765 108865: contig of 2723 bp in length
108865 111347: gap of unknown length
111347 111447: contig of 2482 bp in length
111447 114803: gap of unknown length
114803 114903: contig of 3356 bp in length
114903 116772: gap of unknown length
116772 116872: contig of 1869 bp in length
116872 118340: gap of unknown length
118340 118440: contig of 1468 bp in length
118440 120587: gap of unknown length
120587 120687: contig of 2147 bp in length
120687 122289: gap of unknown length
122289 122390: contig of 1602 bp in length
122390 124086: gap of unknown length
124086 124186: contig of 1697 bp in length
124186 126276: gap of unknown length
126276 126376: contig of 2090 bp in length
126376 127942: gap of unknown length
127942 128042: contig of 1566 bp in length
128042 129220: gap of unknown length
129220 129220: contig of 1178 bp in length

129221 129320: gap of unknown length
129320 130478: contig of 1158 bp in length
130478 130578: gap of unknown length
130578 131805: contig of 1227 bp in length
131805 131905: gap of unknown length
131905 133279: contig of 1374 bp in length
133279 133379: gap of unknown length
133379 135076: contig of 1696 bp in length
135076 135175: gap of unknown length
135175 137317: contig of 2142 bp in length
137317 137417: gap of unknown length
137417 138750: contig of 1333 bp in length
138750 138850: gap of unknown length
138850 141412: contig of 2562 bp in length
141412 141512: gap of unknown length
141512 142664: contig of 1152 bp in length
142664 142764: gap of unknown length
142764 144121: contig of 1357 bp in length
144121 144221: gap of unknown length
144221 146975: contig of 2754 bp in length
146975 147075: gap of unknown length
147075 148195: contig of 1120 bp in length
148195 148295: gap of unknown length
148295 149533: contig of 1238 bp in length
149533 149633: gap of unknown length
149633 151328: contig of 1695 bp in length
151328 151428: gap of unknown length
151428 153745: contig of 2317 bp in length
153745 153845: gap of unknown length
153845 155423: contig of 1578 bp in length
155423 155523: gap of unknown length
155523 157366: contig of 1843 bp in length
157366 157466: gap of unknown length
157466 159157: contig of 1691 bp in length
159157 159257: gap of unknown length
159257 160289: contig of 1032 bp in length
160289 160389: gap of unknown length
160389 161589: contig of 1200 bp in length
161589 161590: gap of unknown length

Query Match 1.5%; Score 21; DB 2; Length 176912;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0;
Gaps 0;
QY 24 cccaagcgacacaagcctt 44
|||||
Db 82779 CCCAAGCGACACAAGCCTT 82799

Search completed: May 22, 2002, 08:46:53
Job time: 11703 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 06:36:01 ; Search time 3530.57 Seconds
(without alignments)
8304.064 Million cell updates/sec

Title: US-09-693-205-7_COPY_5300_6700

Perfect score: 1401
Sequence: 1 gtacgcagcaaacactagagc.....aaagacacttaagaagt 1401

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vl:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.sts:*
27: em.un:*
28: em.vl:*
29: em.un:*
30: em.htg.hum:*
31: em.htg.inv:*
32: em.htg.other:*
33: em.htgo.inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description
------------	-------------	--------	----	-------------

1	1390	99.2	12793	6	AX119931	AX119931 Sequence
2	1390	99.2	12793	9	AF193556	AF193556 Homo sapi
3	1390	99.2	92693	9	AL157766	AL157766 Human DNA
4	1364.6	97.4	99819	2	AC079761	AC079761 Homo sapi
5	1174	83.8	11492	6	AX119933	AX119933 Sequence
6	1174	83.8	11493	10	AF193557	AF193557 Mus muscu
7	126.6	9.0	78844	8	AB06708	AB06708 Arabidops
8	115.2	8.2	165299	2	OSJN0002	AL606457 Oryza sat
9	58	4.1	174140	2	AC069017	AL6065017 Mus muscu
10	54.4	3.9	318221	2	PFMA13P3	AL049184 Plasmodi
11	51.8	3.7	1141	6	AX083744	AL049184 Plasmodi
12	45	3.2	2389	3	AF151112	AF151112 Dictyoste
13	45	3.2	3885	3	AF151733	AF151733 Dictyoste
14	45	3.2	3985	3	AF151111	AF151111 Dictyoste
15	44.6	3.2	160203	2	HS425C14	Z39129 Human DNA s
16	44.4	3.2	114532	2	AC016547	AC016547 Homo sapi
17	44.4	3.2	129046	9	AC008835	AC008835 Homo sapi
18	44.4	3.2	143358	2	AC044843	AC044843 Homo sapi
19	44.4	3.2	161133	2	AC019295	AC019295 Homo sapi
20	44.4	3.2	163690	2	AC083781	AC083781 Homo sapi
21	44	3.1	282928	2	AC091244	AC091244 Rattus no
22	43.4	3.1	187413	9	AC007567	AC007567 Homo sapi
23	43.2	3.1	7218	6	166494	166494 Sequence 14
24	43.2	3.1	83650	8	AB023041	AB023041 Arabidops
25	43.2	3.1	187193	2	AC105657	AC105657 Rattus no
26	43	3.1	231450	1	AL596163	AL596163 Listeria
27	42.6	3.0	169019	2	AC078952	AC078952 Homo sapi
28	42.6	3.0	243675	10	AC079869	AC079869 Mus muscu
29	42	3.0	199890	2	AC095674	AC095674 Rattus no
30	41.6	3.0	102406	2	AC103075	AC103075 Rattus no
31	41.4	3.0	110000	2	AC011600_1	Continuation (2 of
32	40.8	2.9	35338	8	CAC49C4	AL033503 C.albican
33	40.8	2.9	167941	9	AC092638	AC092638 Homo sapi
34	40.8	2.9	234112	3	PFMAL4P2	AL0335475 Plasmodi
35	40.6	2.9	7990	6	AX281265	AX281265 Sequence
36	40.6	2.9	7990	6	AX345060	AX345060 Sequence
37	40.6	2.9	28125	3	U40426	U40426 Caenorhabdl
38	40.4	2.9	50886	9	AL450107	AL450107 Human DNA
39	40.4	2.9	70326	2	AC068957	AC068957 Homo sapi
40	40.4	2.9	161714	2	AC092087	AC092087 Canis fam
41	40.4	2.9	170928	9	AC034148	AC034148 Homo sapi
42	40.4	2.9	190937	9	AP000957	AP000957 Homo sapi
43	40.4	2.9	340000	9	AP001680	AP001680 Homo sapi
44	40.2	2.9	122241	2	AP004129	AP004129 Oryza sat
45	40	2.9	52706	2	AC101794	AC101794 Mus muscu

ALIGNMENTS

RESULT 1
AX119931
LOCUS AX119931 12793 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 1 from Patent WO0129266.
ACCESSION AX119931
VERSION AX119931.1 GI:14036678
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 12793)
Hudson,T.J., Engert,J. and Richter,A.
TITLE Identification of arsacs mutations and methods of use therefor
JOURNAL Patent: WO 0129266-A 1 26-Apr-2001;
MCGILL UNIVERSITY (CA) ; Hopital Sainte-Justine (CA)
FEATURES
Source
1. 12793
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT	ORIGIN
4163 a	2256 c 2487 g 3887 t

Query Match 99.2%; Score 1390; DB 6; Length 12793;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1401; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```

QY 1 gtagcagtaaaactagagcagtcaccaagcagcaacaagccttagaagaatgcattcc 60
DB 5300 GTTAGCAGTAAACTAGAGCAGTCCCAAGCGACACAAACCTTAGAATAATATGATCC 5359
QY 61 aatgctgttttaacaacttgacagaaattggcgagaaaataatgaccagaga 120
DB 5360 AATGCTGTTTAAACAACACTTGGCAGCAATTTGGCGAAGAAAATTTGACCAGAGA 5419
QY 121 attaagaagcattccttaatacatalcctctcgaagaagaattgtgaagaagctcttcaa 180
DB 5420 ATTAAGAGCATCCTTAATGATATCCTCTTGAAGAAGAAATGTTGAAGAAGCTTTCMA 5479
QY 181 aatgctgtagatgacaaaggcgacagaaatctgttctgttctgtagccttagacacatcca 240
DB 5480 AATGCTGATGATCCAAAGCGCAGCAAAATCTGTTTGTGTGATCTTACAGCATCCA 5539
QY 241 gttgtagaataattttagatgaatggcgcccatgtgcaaggcgacagacttgtgtac 300
DB 5540 GTTGATGAAATATTTGATGATAGTGGGCCCATGTGCAAGGCGCAGCATTTGTGTAC 5599
QY 301 aacaacagccatttacaagaagaatgattgaagaatltcagaatcttgaagaagcagc 360
DB 5600 AACACCGACCATTTACAGAGATGATGTTAGAGAAATTCAGAAATCTTGAAAAAGCAGC 5659
QY 361 aaagagggaatccctataaactgaactgacaglatggaatagattcaattctgtatcat 420
DB 5660 AAAGAGGGAATCCTTATTAACCTGACAGTATGGAATAGAAATCTTGATATCAT 5719
QY 421 atcacagactgcccatttattcttggaatgacatccctgttatttggctccat 480
DB 5720 ATCACAGACTGCCATCTTTATTTCTGGCAATGACATCCTGTGATTTTGTGATCCCAT 5779
QY 481 gccacatagcaccaaggcgccacatccatagtcgccgagcagctgtttagaatttgat 540
DB 5780 GCCAGATATGCACACGAGGCGCATCATATAGTCCGAGAGCATGTTTAGAATTTGAT 5839
QY 541 gcagattttagagacacagttctcagatgtctgagatcttattctgggaaccattttaa 600
DB 5840 GCAGATTTTAAAGACAGATCTTCAGATGTTCTGATCTTATCTTGGGCAACCCATTTTAA 5899
QY 601 ctgagtaattgacacaatgttccagattcctctcgttaatgcagaatggcaaaagtctgc 660
DB 5900 CTGATTAATTTGCACAAATGTTCAATTTCTCTGTAATGCAGAAATGGCAAAAGTTTCG 5959
QY 661 gaaattctgtcttccagacatcagacagaatgtccagaatcttltgacaactgcgc 720
DB 5960 GAAATTTTCGTCGTTCAGATCAGACAGAAATGTCAGAAATCTTTTGGCAAACTGCGC 6019
QY 721 tcgagatggggcagaactcttaagtcttcaatcacatgtaaaaaattctattgtgaa 780
DB 6020 TCAGATGGGGCAGAACTTCTTAATGTTCTTAAATCACATGAAAAAATTTCTATTGTGAA 6079
QY 781 atagataagagtagttagctcgaatgtgctgatttcagtaaaggcacaatcacagat 840
DB 6080 ATAGATAAGAGTAGTGGAGCTTAATGTCTTATTTCACTAAGGCAAAATCACAGAT 6139
QY 841 ggaagacagattgaaaaagaacaattctcatgcatctgtaattgatagttactaaaga 900
DB 6140 GGAGACAGATTTGAAAAGGAACAATTTCAATGCATGTAATGATAGTGTACTAAAAAG 6199
QY 901 aggcagctcaagaacataccagttcaacaataactatactatgtagtactgagactct 960
DB 6200 AGGCAGCTCAAAACACATACAGTTCACAAATTAACCTATATGATGATCTGAGGACTCT 6259
QY 961 gaaggaatcttactcagtggtcgaatttgaatagatcaggtcttccaagtatggagaa 1020
DB 6260 GAAGGAATCTTACTACGTGCTAATTTGTAAATAGATCAGGCTTTTCAAGTATGAGAAA 6319

```

```

QY 1021 gtacltaaaagltcatatcagctcacaaagaacaaagatalactcttctccacgttgt 1080
DB 6320 GTATCTTAAAGGTGCATATCAGCTCACAAAGAACAAATATTAATCTTTTCCACGTGT 6379
QY 1081 ggaatagctgctgcattactcacaaactataaaaaacccatagggtctctgtttttg 1140
DB 6380 GGAATAGCTCTCCGTCACTACTACAACTATTAATAAACCCATAGGGCTTCTGTGTTTGG 6439
QY 1141 cctcttcttcttggaactggctgcacattcattgtagaattggcacttggacttgatcca 1200
DB 6440 CCTCTTTCTTTGGAGACTGGGCTGCCATTTTCATGTGAATGGCCACTTTGCACTGGATTCA 6499
QY 1201 gccagaagaacctgtgtgcgtgataataatgagttgtgttcgaagtgaactggaataac 1260
DB 6500 GCCAAGAGGAACCTGTGGCTGATGATTAATGGAATGCTGTGTGAAGTGAAGTGAATTAAC 6559
QY 1261 agtttaatgacacatataatagctcctgcatatg-tyaattgtcaatacagttaaaaaa 1319
DB 6560 AGTTTAATGACACACTTAATAGCTCCGTGCAATATGTTGAATGCTTAATGCTTAATGCTTA 6619
QY 1320 cggtaattccctgtgtctgtagtccacatatacagttgttaccagaacaccctattcatgt 1379
DB 6620 CGGTATTTCCCTGGTGTGTGATTCACACTTATTCAGTGTTAACAGAACCCCTTATTCATGTT 6679
QY 1380 gtaaaagacactttaagaagt 1401
DB 6680 GTAAGGACACTTTAAAGAAGT 6701

```

```

RESULT 2
AF193556 12793 bp DNA linear PRI 07-FEB-2000
LOCUS Homo sapiens saccin (SACS) gene, complete cds.
DEFINITION AF193556
ACCESSION AF193556
VERSION AF193556.1 GI:6907041
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 12793)
Engert,J.C., Berube,P., Mercier,J.J., Dore,C., Lepage,P., Ge,B.,
Bouchard,J.P., Mathieu,J., Melancon,S.B., Schalling,M.,
Lander,E.S., Morgan,K., Hudson,T.J. and Richter,A.
ARSACS, a spastic ataxia common in northeastern Quebec, is caused
by mutations in a new gene encoding an 11.5-kb ORF
Nat. Genet. 24 (2), 120-125 (2000)
JOURNAL 20120709
MEDLINE 2 (bases 1 to 12793)
REFERENCE Engert,J.C., Berube,P., Dore,C., Lepage,P., Ge,B., Hudson,T.J. and
AUTHORS Richter,A.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-1999) Genome Centre, Montreal General Hospital,
1650 Cedar Ave., Montreal, QC H3G 1A4, Canada
FEATURES
source
1..12793
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="between D13S232 and D13S292"
1..12793
/gene="SACS"
/product="sacsln"
1..12793
/gene="SACS"
77..11566
/feature="SACS"
/note="molecular chaperone"
/codon_start=1
/product="sacsln"
/protein_id="AF31262.1"
/db_xref="GI:6907042"
/translation="MNTFWPGRELLIVQWYPFDENRNHPSVSLKMYKMLYIHSEDL

```


QY 1081 ggaagtagctgcctgacttactcaacaataaaaaacccatagggccttcgtttttg 1140
 |||||
 Db 6380 GAGGTAGCTCCCTGCATCTACTCACAATAAAAAACCCATAGGCCCTTCTGTTTTG 6439
 |||||
 QY 1141 cctcttcttggagagctggctgctcattcattgtaagtgccactttgactgattca 1200
 |||||
 Db 6440 CCTCTTCTTGGAGAGACTGGCTGCCATTTTCATGTGAATGGCCACTTTCGACTGATTC 6499
 |||||
 QY 1201 gccagaagagactctggtggtgtagtaataatgagtggtgttttgcgaagtactggaataac 1260
 |||||
 Db 6500 GCCAGAGAGAACTGTGGCTGATGATATGAGAGTTGGTTCGAAGTGAAGTGAATAC 6559
 |||||
 QY 1261 agttaatgacagcatatagctcctgcatatg-tgaattgctaatacagttaaaaaa 1319
 |||||
 Db 6560 AGTTTATGACACATTAATAGCTCCGATATGTTGAATTGCTAATACAGTTAAAAAA 6619
 |||||
 QY 1320 cggatttcccttgcttgcataacatcagtggttaacagaacacctatcatgtt 1379
 |||||
 Db 6620 CGGTATTTCCCTGTTCTGATCCACATTTATCAGTTACAGAACACCCCTATTCATGTT 6679
 |||||
 QY 1380 gtaaaagcaccttaagaagt 1401
 |||||
 Db 6680 GTAAAGCACCTTTAAAGAGT 6701
 |||||

RESULT 3
 AL157766/c 92693 bp DNA linear PRI 11-APR-2001
 LOCUS Human DNA sequence from clone RP11-40020 on chromosome
 DEFINITION 13q12.11-12.2, complete sequence.
 ACCESSION AL157766
 VERSION AL157766.9 GI:13620292
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 92693)
 TROMANS, A.
 Direct Submission
 Submitted (11-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 Requests: clonerequest@sanger.ac.uk
 On Apr 12, 2001 this sequence version replaced gi:12709868.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit submissions with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30);
 an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em, EMBL; Sw,
 SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP
 database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 13, constructed by the Sanger Centre Chromosome 13
 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr13
 RP11-40020 is from the library RPCI-11.1 constructed by the group
 of Pieter de Jong. For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pBACE3.6
 IMPORTANT: This sequence is not the entire insert of clone
 RP11-40020. It may be shorter because we sequence overlapping
 sections only once, except for a 100 base overlap.

FEATURES
 source The true left end of clone RP11-760M1 is at 92594 in this sequence.
 The true right end of clone RP11-72P19 is at 100 in this sequence.
 Location/Qualifiers

1..92693
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="13"
 /map="q12.11-12.2"
 /clone="RP11-40020"
 /clone_11b="RPCI-11.1"
 2390..2485
 /note="MIR repeat: matches 81..192 of consensus"
 2562..2673
 /note="MIR repeat: matches 68..183 of consensus"
 3896..4201
 /note="AluY repeat: matches 3..308 of consensus"
 5122..5397
 /note="MER46C repeat: matches 1..286 of consensus"
 18986..19294
 /note="AluY repeat: matches 1..310 of consensus"
 19644..19873
 /note="MER46A repeat: matches 1..235 of consensus"
 20613..20912
 /note="AluSg1 repeat: matches 1..300 of consensus"
 23342..23651
 /note="AluSg1 repeat: matches 1..309 of consensus"
 24769..24891
 /note="L2 repeat: matches 2554..2682 of consensus"
 23871..26011
 /note="L2 repeat: matches 2356..2495 of consensus"
 26033..26109
 /note="L2 repeat: matches 2601..2688 of consensus"
 26245..26344
 /note="L2 repeat: matches 2154..2255 of consensus"
 2638..27096
 /note="MIR repeat: matches 3..175 of consensus"
 27150..27653
 /note="L2 repeat: matches 1063..1644 of consensus"
 28522..28891
 /note="TMR1B repeat: matches 1..364 of consensus"
 29447..29834
 /note="L1ME3A repeat: matches 5787..6164 of consensus"
 36098..36415
 /note="AluSx repeat: matches 1..308 of consensus"
 37202..37414
 /note="MIR repeat: matches 22..262 of consensus"
 37963..38254
 /note="AluSg repeat: matches 9..301 of consensus"
 38703..39008
 /note="AluSg repeat: matches 1..306 of consensus"
 39790..40093
 /note="AluSx repeat: matches 1..304 of consensus"
 40126..40416
 /note="AluSg repeat: matches 1..292 of consensus"
 40444..40733
 /note="AluSg repeat: matches 1..292 of consensus"
 41322..41405
 /note="Single clone region. Assembly confirmed by
 restriction digest data."
 41541..41788
 /note="AluSg repeat: matches 1..248 of consensus"
 44790..45101
 /note="AluSg repeat: matches 1..313 of consensus"
 45261..45312
 /note="13 copies 4 mer tggc 888 conserved"
 45899..46206
 /note="AluY repeat: matches 1..307 of consensus"
 46754..47052
 /note="AluY repeat: matches 1..298 of consensus"
 47067..47365
 /note="AluY repeat: matches 1..299 of consensus"
 47477..47873
 /note="L1MA10 repeat: matches 5950..6322 of consensus"

```
repeat_region 47889..48229
                /note="AlusX repeat: matches 1..312 of consensus"
repeat_region 49168..49212
                /note="Alu repeat: matches 85..126 of consensus"
repeat_region 49620..49693
                /note="L2 repeat: matches 1685..1757 of consensus"
repeat_region 50704..51032
                /note="AlusC repeat: matches 1..308 of consensus"
misc_feature 52204..53009
                /note="CpG Island"
                /evidence="not_experimental"
repeat_region 53978..54137
                /note="AlusX repeat: matches 1..160 of consensus"
repeat_region 54179..54511
                /note="LMB6 repeat: matches 5822..6172 of consensus"
repeat_region 53685..55949
                /note="AlusX repeat: matches 1..305 of consensus"
repeat_region 57331..57390
                /note="AlusX repeat: matches 1..305 of consensus"
repeat_region 57357..57392
                /note="30 copies 2 mer ga 75% conserved"
repeat_region 57753..57930
                /note="9 copies 4 mer gaga 91% conserved"
repeat_region 58260..58389
                /note="MIR repeat: matches 82..262 of consensus"
repeat_region 58564..58611
                /note="MIR repeat: matches 2..153 of consensus"
repeat_region 59350..59533
                /note="24 copies 2 mer ca 93% conserved"
repeat_region 59992..60223
                /note="AlusQ repeat: matches 129..313 of consensus"
repeat_region 61036..61144
                /note="AluJo repeat: matches 85..299 of consensus"
repeat_region 62008..62187
                /note="L2 repeat: matches 2581..2656 of consensus"
repeat_region 62188..62316
                /note="TIGER1 repeat: matches 2238..2418 of consensus"
repeat_region 62330..62363
                /note="AluY repeat: matches 1..129 of consensus"
repeat_region 62362..62565
                /note="Alu repeat: matches 261..294 of consensus"
repeat_region 62566..62865
                /note="TIGER1 repeat: matches 1586..1787 of consensus"
repeat_region 62866..64385
                /note="AlusQ repeat: matches 1..302 of consensus"
repeat_region 64386..64694
                /note="TIGER1 repeat: matches 46..1586 of consensus"
repeat_region 64695..64713
                /note="AluY repeat: matches 1..306 of consensus"
repeat_region 65068..65395
                /note="TIGER1 repeat: matches 29..46 of consensus"
repeat_region 65396..65569
                /note="LIPB2 repeat: matches 5405..5733 of consensus"
repeat_region 65571..65640
                /note="AluY repeat: matches 136..309 of consensus"
repeat_region 65696..65717
                /note="LIPB2 repeat: matches 5728..5791 of consensus"
repeat_region 65725..66096
                /note="11 copies 2 mer ta 100% conserved"
repeat_region 66371..66410
                /note="LIPB2 repeat: matches 5789..6155 of consensus"
repeat_region 67586..67886
                /note="10 copies 4 mer tgtg 82% conserved"
repeat_region 69748..69930
                /note="Aluub repeat: matches 1..299 of consensus"
repeat_region 70957..71267
                /note="MIR repeat: matches 6..248 of consensus"
repeat_region 71279..71413
                /note="AluY repeat: matches 1..311 of consensus"
repeat_region 71411..71737
                /note="MER21B repeat: matches 548..680 of consensus"
repeat_region 71780..72075
                /note="MER31a repeat: matches 47..485 of consensus"
repeat_region 71780..72075
                /note="AlusX repeat: matches 1..295 of consensus"
```

```
repeat_region 72145..72256
                /note="MER31-internal repeat: matches 42..175 of
                consensus"
repeat_region 72454..72865
                /note="MER31-internal repeat: matches 332..739 of
                consensus"
repeat_region 72873..73249
                /note="MER31-internal repeat: matches 883..1261 of
                consensus"

Query Match 99.2% Score 1390; DB 9; Length 92693;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1401; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 gtagcaataaactcaggagcagctcccaagcgacacaaagccttagaagaatgcatcc 60
DB 13610 GTAGCACTAAACCTAGAGCGAGCTCCCAAGGACACAAAGCCTTAGAAGATTAGCATCC 13551

QY 61 aatgctgtttacaacacctgycacagaatttggcagaagaagaattaccagcaga 120
DB 13550 AATGCTGTTTACACACACTTGGCAGCAAAATTGGCAGAAAGAAAATTACCAGCAGA 13491

QY 121 attaagacatcccttaatgcatatcccttcgaaaagaatgtaagaagctcttcoa 180
DB 13490 ATTAAGACATCCTTAATGCATATCTTGAAAAGAAAATTGAAAAGACCTTCTTCAA 13431

QY 181 aatgctgatgcaaaagcgacagaaatctggttggttgttgatctctagaacatcca 240
DB 13430 AATGCTGATGCAAAAGCGACAGAAATCGTTTGTGTTGATCTTACACATCCA 13371

QY 241 gttgataaatattgtagtaagtggcccatctgcaagggcgacacattgtgttac 300
DB 13370 GTTAGTAATATTGATGATAGTAAGTGGCCCATATTGCAAGGCCACACTTGTGTATC 13311

QY 301 aacacaccagccattacagaagaatgcatgtaagaagaaattcgaatcttggaaaagcagc 360
DB 13310 AACACACAGCCATTACAGAAAGATGATGTAGAGAAATTCAGAAATTCGAAAAGCAGC 13251

QY 361 aaagaaggaaatccctataaacctgacagatgtaagaatgtaaatccatctgtatcat 420
DB 13250 AAAGAGGAAATCCTTATTAACAGAGATGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTA 13191

QY 421 atcacagactgcccatttatttattcttcgcaatgacatccggtgatatttgatcccat 480
DB 13190 ATCACAGACTGCCATCTTTATTCTTGCAATACATCCTGTATTTTGTATCTCAT 13131

QY 481 gccagatatgcaccaggggcacacatccatagctccggaagcagatgtaagaattgcat 540
DB 13130 GCCAGATATGCACAGGGGCCACATCCATTAGTCCGAGCCAGATTAGAGATTGGAT 13071

QY 541 gcagattttagcacagcttcccaatgcttgcagatcttcttcggaaccatcttaaa 600
DB 13070 GCAGATTTTAGCACAGCTTCTTCAAGATGCTTGTATCTTATTCGGAACCATTTTAA 13011

QY 601 ctgataatgcaaatgcttcaagattcccttccttgtaagtgaagaatggaagaatttcg 660
DB 13010 CTGATATTTGCACAAATGTTTCAAGATTCTTCTTGTATGCAAGAAATGCAAAAGTTTCG 12951

QY 661 gaaattcgtctgttcagcatcagacagaatggtccagaatcttcttggacaaactgcgc 720
DB 12950 GAAATTTGCTGTTCACAGCATCAGACAGAAATGTCAGAAATCTTTGGACAAACTGCGC 12891

QY 721 tcagatgggacagaactcttaatgcttcttaatacatatgtaaaatcttcaatttgtaa 780
DB 12890 TCAGATGGGACAGAACTTCAATGTTTCTTAATACATGGAATAAATTTCAATTTGTGAA 12831

QY 781 atagataaagtaactgagactctaaatgctgctgattacagtaaaaggcaaatcacagat 840
DB 12830 ATAGATTAAGACTGAGACTCTTAATGCTGTGATTTCAGTAAAGGCAAAATACAGAT 12771

QY 841 ggaagacagattgaaaagaaacaaattcatgcatctgtaattgtaagtgtactaaaag 900
DB 12770 GGACACAGATTGAAAAGAAAGAAATTCATGCAATCTGTAATGTAGATGTTACTAAAG 12711
```

QY	901	aggagcgcacaagacatccagcttcaacaataaaccttactgtgatactggaacct	960
Db	12710	AGCGAGCTCAAGACATACCACTTACAAATAATACCTATACATGATACGACACTCT	12651
QY	961	gaaggaatcttactacgtggtctaatttgttaataagatcaggtcttcaagtaaggaaa	1020
Db	12650	GAAGGAATCTTACTACCTGGCTTAATTTGTAATAGATCAGGCTTTCAAGATGAGAAA	12591
QY	1021	gtatctaaagtgtcatatcagctctcagaagcaagatatctctttcccaagtgtg	1080
Db	12590	GTATCTAAAGTGTATATACGCTACAGAACCAAGANTATCTTTTCCACGTGGT	12531
QY	1081	ggaagtgcctgcctgattacttacaacaataaaaaaacccataaggccctctgtcttttg	1140
Db	12530	GGAGTAGCTGCCTGCATTAACACATATATAAAAAACCCATAGAGGCCCTTGCTTTTGG	12471
QY	1141	cctcttcttttggagacctgggtcgcatttcaatgtgaatgaccttgcactgtattca	1200
Db	12470	CCCTCTTTCTTGGAGACACGGGCGCTGCCAATTTCTATGGAATGGCCACTTTGCACTGGATTCA	12411
QY	1201	gccagaagaacctgtggtcgatgatatatgagttgtgttcggaagtgcacgtggaataac	1260
Db	12410	GCCAGAGAGAACCTTGGCGGTATATATATGACAGTGGTGTTCCAGAGTACTGGAAATATAC	12351
QY	1261	agttaatgacagcaatataatgctctgtcataatg- tgaattgtcataatacagttaaaaaa	1319
Db	12350	AGTTAAATGACAGCATTTAATAGCTCTGCATATGTTGAATTTCTAATACAGTTAAAAAAA	12291
QY	1320	cggatcttcctgtgttcgataccaacatatacagtggtacagaagaaacccctattatgt	1379
Db	12290	CGGATTTTCCTCGTTCGATTCACCACTATATCAGTGTTCAGAAACACCCCTATTTATGTT	12231
QY	1380	gtaaagacactttaagaagt 1401	
Db	12230	GTAAGGACACTTTTAAAGAGT 12209	
RESULT 4	AC079761/c	99819 bp DNA linear HTG 10-SEP-2000	
LOCUS	AC079761		
DEFINITION	Homo sapiens chromosome YUK clone RP11-143617, *** SEQUENCING IN		
ACCESSION	AC079761		
VERSION	AC079761.1	GI:10047966	
KEYWORDS	HTG; HTGS; PHASE1.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Mumaiyola; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Manualla; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 99819)		
JOURNAL	Waterston, R.H.		
REFERENCE	2 (bases 1 to 99819)		
AUTHORS	Waterston, R.H.		
TITLE	Direct Submission		
JOURNAL	Submitted (10-SEP-2000) Genome Sequencing Center, Washington		
	University School of Medicine, 4444 Forest Park Parkway, St. Louis,		
	MO 63108, USA		
COMMENT			
	----- Genome Center -----		
	Center: Washington University Genome Sequencing Center		
	Web site: http://genome.wustl.edu/gsc/index.shtml		
	Project Information -----		

	* NOTE: This is a 'working draft' sequence. It currently		
	* consists of 44 contigs. The true order of the pieces		
	* is not known and their order in this sequence record is		
	* arbitrary. Gaps between the contigs are represented as		
	* runs of N, but the exact sizes of the gaps are unknown.		
	* This record will be updated with the finished sequence		
	* as soon as it is available and the accession number will		
	* be preserved.		

1	1137	contig of 1137 bp in length
1138	1237	gap of unknown length
1238	2558	contig of 1301 bp in length
2539	2638	gap of unknown length
2639	3976	contig of 1338 bp in length
3977	4077	gap of unknown length
4077	5355	contig of 1279 bp in length
5356	5455	gap of unknown length
5456	6757	contig of 1302 bp in length
6758	6857	gap of unknown length
6858	8570	contig of 1713 bp in length
8571	9935	contig of 1255 bp in length
9926	10025	gap of unknown length
10026	11426	contig of 1401 bp in length
11427	11526	gap of unknown length
11527	13286	contig of 1740 bp in length
13287	13367	gap of unknown length
13367	14794	contig of 1428 bp in length
14795	14894	gap of unknown length
14895	16054	contig of 1160 bp in length
16055	16154	gap of unknown length
16155	17395	contig of 1241 bp in length
17396	17495	gap of unknown length
17496	19287	contig of 1792 bp in length
19288	19387	gap of unknown length
19388	21294	contig of 1907 bp in length
21295	21394	gap of unknown length
21395	22944	contig of 1550 bp in length
22945	23044	gap of unknown length
23045	24421	contig of 1377 bp in length
24422	24521	gap of unknown length
24522	25870	contig of 1349 bp in length
25871	25970	gap of unknown length
25971	27230	contig of 1260 bp in length
27231	27330	gap of unknown length
27331	27738	contig of 1448 bp in length
28779	28878	gap of unknown length
28879	30893	contig of 2015 bp in length
30894	30993	gap of unknown length
30994	32460	contig of 1467 bp in length
32461	32560	gap of unknown length
32561	33984	contig of 1424 bp in length
33985	34084	gap of unknown length
34085	35285	contig of 1201 bp in length
35286	35385	gap of unknown length
35386	37184	contig of 1799 bp in length
37185	37284	gap of unknown length
37285	39172	contig of 1888 bp in length
39173	39272	gap of unknown length
39273	40874	contig of 1602 bp in length
40875	40974	gap of unknown length
40975	42893	contig of 1919 bp in length
42894	42993	gap of unknown length
42994	44384	contig of 1391 bp in length
44385	44484	gap of unknown length
44485	45999	contig of 1515 bp in length
46000	46099	gap of unknown length
46100	48699	contig of 2570 bp in length
48700	48796	gap of unknown length
48770	50798	contig of 2029 bp in length
50799	50898	gap of unknown length
50899	52809	contig of 1911 bp in length
52810	52909	gap of unknown length
52910	55127	contig of 2218 bp in length
55128	55227	gap of unknown length
55228	58087	contig of 2860 bp in length
58088	58187	gap of unknown length
58188	61004	contig of 2817 bp in length
61005	61104	gap of unknown length
61105	64185	contig of 3061 bp in length
64186	64285	gap of unknown length
64286	67105	contig of 2820 bp in length

	*	67106	67205: gap of unknown length	/note="assembly_name:Contig65"
	*	67206	70837: contig of 3632 bp in length	42994. .44384
	*	70838	70937: gap of unknown length	/note="assembly_name:Contig66"
	*	70938	75837: contig of 4900 bp in length	44485. .45999
	*	75838	75937: gap of unknown length	/note="assembly_name:Contig67"
	*	75938	80452: contig of 4515 bp in length	46100. .48669
	*	80453	80552: gap of unknown length	/note="assembly_name:Contig68"
	*	80553	84661: contig of 4109 bp in length	48770. .50798
	*	84662	90542: contig of 5781 bp in length	/note="assembly_name:Contig69"
	*	90543	90642: gap of unknown length	50899. .52809
	*	90643	94348: contig of 3706 bp in length	/note="assembly_name:Contig70"
	*	94349	94448: gap of unknown length	52910. .55127
	*	94449	99819: contig of 5371 bp in length.	/note="assembly_name:Contig71"
FEATURES			Location/Qualifiers	
SOURCE			1. .99819	
			/organism="Homo sapiens"	
			/db_xref="taxon:9606"	
			/chromosome="UNK"	
			/clone="RP11-143G17"	
			1. .1137	
misc_feature			/note="assembly_name:Contig20"	
misc_feature		1238. .2358	/note="assembly_name:Contig23"	
misc_feature		2639. .3976	/note="assembly_name:Contig30"	
misc_feature		4077. .5355	/note="assembly_name:Contig35"	
misc_feature		5456. .6757	/note="assembly_name:Contig39"	
misc_feature		6858. .8570	/note="assembly_name:Contig40"	
misc_feature		8671. .9925	/note="assembly_name:Contig41"	
misc_feature		10026. .11426	/note="assembly_name:Contig42"	
misc_feature		11527. .13266	/note="assembly_name:Contig44"	
misc_feature		13367. .14794	/note="assembly_name:Contig47"	
misc_feature		14895. .16054	/note="assembly_name:Contig48"	
misc_feature		16155. .17395	/note="assembly_name:Contig49"	
misc_feature		17496. .19287	/note="assembly_name:Contig50"	
misc_feature		19388. .21294	/note="assembly_name:Contig51"	
misc_feature		21395. .22944	/note="assembly_name:Contig52"	
misc_feature		23045. .24421	/note="assembly_name:Contig53"	
misc_feature		24522. .25870	/note="assembly_name:Contig54"	
misc_feature		25971. .27230	/note="assembly_name:Contig55"	
misc_feature		27331. .28778	/note="assembly_name:Contig57"	
misc_feature		28879. .30893	/note="assembly_name:Contig58"	
misc_feature		30994. .32460	/note="assembly_name:Contig59"	
misc_feature		32561. .33984	/note="assembly_name:Contig60"	
misc_feature		34085. .35285	/note="assembly_name:Contig61"	
misc_feature		35386. .37184	/note="assembly_name:Contig62"	
misc_feature		37285. .39172	/note="assembly_name:Contig63"	
misc_feature		39273. .40874	/note="assembly_name:Contig64"	
misc_feature		40975. .42893		

Query Match	97.48;	Score 1364.6;	DB 2;	Length 99819;
Best Local Similarity	99.28;	Pred. No. 1.5e-309;		
Matches 1392;	Conservative	0;	Mismatches 9;	Indels 2;
			Gaps	2;
QY	1	gtacagtaaaactcagagcagctcccaagcgacacaaagcctttagaagaatgcatcc	60	
Db	88196	GTACAGTAANAACCTTAGGAGCGAGTCCTCCANAGCACAAAGCCTTAGAAGATATGCATCC	88137	
QY	61	aatgctgttttaaacacaccttgacagaaatttggcagaaagaaatltgacagcaga	120	
Db	88136	AATGCTGTTTACAAACCTTGACACAAATTTGGCGAAGAAATGACCAGCAGA	88077	
QY	121	attaagagcatccttaatgatcatccctctgaaaagaatgttgaagcttctca	180	
Db	88076	ATTAAAGACATCCTTAAGCATATCTTCTGAAAAGGAATGTGAAGACCTTCTTCAA	88017	
QY	181	aatcctgatagtgcaaaagcgacagaaatctgttttggttgatccctagaacatcca	240	
Db	88016	AATGCTGATGTGCAAAAGCGACAGAAATCTGTTTGTTGTTGATCCTAGACAGCATCCA	87957	
QY	241	gttgataaatatttgatgataaagtggcccccatctgcaaggcgacacattgtgtac	300	
Db	87956	GTTGATTAATATTTGATGATAAGTGGCCCCCATTCGAAGGCCACACTTGTGTATAC	87897	
QY	301	aacaacagccatttacaagaatgtagttagaggaatcagaatcttggaaaagcag	360	
Db	87896	AACAACCGCCATTACAGAGATGATCTTAGAGAAATTCAGAAATCTTGAAGAGCGACG	87837	
QY	361	aaagagggaataccttataaaaactggacagatagaatagaattcaattctgtatcat	420	
Db	87836	AAAGAGGGAATTCCTTATAAACAAGTGACAGATAGGAATGAAATTCATGATATCAT	87777	
QY	421	atcacagactgccatcttatttcttgcaatgacatccggtgtatatttgatcccat	480	
Db	87776	ATCACAGACTGCCCATCTTATTTCGCAATGACATCGTGTATTTTGATTCCTCAT	87717	
QY	481	gccagatatgcacagggggccacatcattagtcctcggagcagtgcttaagattggat	540	
Db	87716	GCCAGATATGCACAGGGGCCACATCCATAGTCCCGCACCATGTTAGAGATTTGGAT	87657	
QY	541	gcagaattttaggacacagtttcagagtcttgatctttagatctttaggggaaccatttaa	600	
Db	87656	GCAGATTTTAGGACACAACTTCTCAGATCTTGTGATCTTTATCTGGGAACCATTTTAA	87597	
QY	601	ctggaataatgcacaatgttccagatttcctcttcgtcaatgcagaataagtcaaaagttcg	660	
Db	87596	CTGGAATAATGCACAATGTTCCAGATTTCTCTGATATGCAGAAATGCGAAAGTTTCG	87537	
QY	661	gaatttcgtctgttcacagcatcagacagaatgttccagaatcttttggacaaactgcgc	720	
Db	87536	GAAATTCGTCTGTTCACGATCAGACAGAAATGTGTCGAAATCTTTTGACAAACTGCGC	87477	

QY 721 tcagatgaggcagaactctcaatgttcttaacacatggaataattctattgtgaa 780
|||||
Db 87476 TCAGATGGGGCAGACACTTAAAGTTTCTTAATCACATGGAATAAATTCTATTGTGAA 87417
QY 781 atagataagatctcagagctctaataatgctglatcagtaaaaggcaaatcacagat 840
|||||
Db 87416 ATGATTAAGAGTCTGAGCTTAAATGTGCTATTCATTAAGGCAAAATCACAGAT 87357
QY 841 ggaagacagattgaaagaaacaaattctcagcctgtaattgatatgttactaaag 900
|||||
Db 87356 GGAAGACGATTGAAAAGGAACAAATTTCATGCAATCTGTAAATTAAGTGTACTAAAAAG 87297
QY 901 aggcagctcaaaagacataccagcttcaacaataaccatactatgaaagagcctc 960
|||||
Db 87296 AGGCAGCTCAAAAGACATACAGATATCAAAATAGCCATATCTATGATTAAGAGCTCT 87237
QY 961 gaaagaaatc-ttaactagctgcttaattgttaataagtcagagcttttcaagtaatgagaa 1019
|||||
Db 87236 GAAAGAAATGTAACATACGCTGAGCTAATTTGTATAGATCAGGCTTCAAGATGAGAA 87177
QY 1020 agtatctaaagtgatcatatcagctcaacaaagaaacaaatatactctttcccaagtg 1079
|||||
Db 87176 AGTATCTAAAAGTGTATATACAGCTCACAAAGAACAAAGATATTACTCTTTTCACAGTGG 87117
QY 1080 tggagtagctgctgcatctactcaacaactataaaacccatagggcctctgctttt 1139
|||||
Db 87116 TGGAGTAGCTGCTGCTGCTATCTCACAACTATTAATAAACCCCATAGGCTCTTCTTTT 87057
QY 1140 gctcttctcttgagagctgagctgctcattcattgtaattgagcgaacttgcactgagttc 1199
|||||
Db 87056 GCCTCTTCTTTTGGAGAGCTGGGCTGCATTTCAATGTAAGGCCACTTTTCACATGAGATTC 86997
QY 1200 agccagaagaaacacttgctgagtgatgataatggaattggttcgaagtagcagagataa 1259
|||||
Db 86996 AGCCAGAAAGAACCTGTGGGTATGTAATGAAGTGGTGGTGAAGTACAGTGAAGTAA 86937
QY 1260 cagcttaatagacagcatlaaagctcctgcatatg-tgaattgctaatacagtaataaa 1318
|||||
Db 86936 CAGTTTAATAGACAGATTATAGCTCTGATATGTTGATTTGATTAATAGTTAAAAA 86877
QY 1319 acggtattccctggtgtctgtaacccaacattacagtgtaacagaacacccattcaatgt 1378
|||||
Db 86876 ACGGTATTTTCCCTGGTGTGATCCAAATATCATGATGTTACAGAACACCCCTATTTCATGT 86817
QY 1379 tgtaaaagacactttaaagaagt 1401
|||||
Db 86816 TGTAAAGGACACTTTAAAGAGT 86794

RESULT 5
AX119933 11492 bp DNA linear PAT 11-MAY-2001
LOCUS AX119933
DEFINITION Sequence 3 from Patent WO0129266.
ACCESSION AX119933
VERSION AX119933.1 GI:14036679
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 11492)
AUTHORS Hudson,T.J., Engert,J. and Richter,A.
TITLE Identification of aarscs mutations and methods of use therefor
JOURNAL Patent: WO 0129266-A 3 26-APR-2001;
MCGILL UNIVERSITY (CA); Hopital Sainte-Justine (CA)
FEATURES
source 1..11492
/organism="Mus musculus"
/db_xref="taxon:10090"

BASE COUNT 3599 a 2280 c 2387 g 3226 t
ORIGIN

Query Match 83.8%; Score 1174; DB 6; Length 11492;
Best Local Similarity 90.3%; Pred. No. 8.8e-265;
Matches 1266; Conservative 0; Mismatches 135; Indels 1; Gaps 1;
QY 1 gtagcagtaaaactaggaagcagtcaccaagcgacacaaagcccttagaaagatgcatcc 60
|||||
Db 5223 GTAGCTGTAAAACTGGTGCATATACCAAGACATTAAGCATTAAGATATGATCATCC 5282
QY 61 aatgctgttttcaacacacttgcagacgaattggcgagaaaataaattgaccacaaga 120
|||||
Db 5283 AACATCTGTTTACAGCTCTAGGTACAGAAATTTGGCGAGAAAGAAACTGACCACAGAG 5342
QY 121 atcaagagcaltccatlaatgcatatccctctgaaaagaaatgltgaaaagagcttccaa 180
|||||
Db 5343 ATTAAGAGCATTTCTCATATGCTTCTTCCATGAAAAGAAATGCTGAAGAGCTTCTTCAA 5402
QY 181 aatgctgataatgcaaaagcgacagaaatcgtttgtgttgcattccctagacagcatcaa 240
|||||
Db 5403 AATGCTGATGATGCAAAAGCCACAGAGATCTGCTTGTGTGATCTGATGACAGCATCTCT 5462
QY 241 gttgataagaatatttgatgataagtggtgccccttgcagaagggccagacttgtgtac 300
|||||
Db 5463 GTTGACCGAATATTTGATGATTAAGTGGGCCCTGCAAGGGCCACACTGTGTGTTTAC 5522
QY 301 aacaaccagccatttacaagagatgataatgtaagaaatcagaatccttggaaaaggcag 360
|||||
Db 5523 AACACCCAGCATTTTACAGAGATGATGTTAGAGAAATTCAGAAATCTTGGAAGAGCCACC 5582
QY 361 aaagaggaaaatccctataaaactgagacaglatgaaatagatcaattcgtgtatcat 420
|||||
Db 5583 AAAGAAAGGGAATCTTGCAAAACAGGACATTAAGATGATGATGATGATGATGATGATGAT 5642
QY 421 ataccagagccacatcttatttcttgcaagaatccctgtgtatttttggctccat 480
|||||
Db 5643 ATTAACAGACTGCCCTCTTTTATTTCTGCAATGATGATGATGATGATGATGATGATGATGAT 5702
QY 481 gccagatatcacacagggcgacacatccatagtcctccggagcagcatgtttagaattgagat 540
|||||
Db 5703 GCCAGATATTCACACAGAGCCACATCATGATGATGATGATGATGATGATGATGATGATGAT 5762
QY 541 gcaagatttagaacaagttcagatgctcagatccttattcttggaaccattttaa 600
|||||
Db 5763 GCAGACTTTTGAACCCAGCTCTCAGATGTTCTGATCTGATCTGATCTGATCTGATCTGATCT 5822
QY 601 ctgataatagcacaatgctcagatttccctcctcgtgaatgcagaatggaagaagtttg 660
|||||
Db 5823 CTGGACAAATGTACAAATGTTTAAATTTCTCTGTAATGCAAGATGGCACAAATTTTCA 5882
QY 661 gaatttcgtctgttcagcatcagacagaaatggtccagatccttcttgacaacatgctgc 720
|||||
Db 5883 GAAATTTCTTCGTTCCATCATATCAGACAGAAATGCTCCAGAAATCTTTGGCAAGTTTACG 5942
QY 721 tcagatgaggcagaactctcaatgttcttaacacatggaataattctattgtgaa 780
|||||
Db 5943 TCTGATGGGGCAGAACCTTAAATGTTTCTCAACACATGAGAAATATCTATTGTGAA 6002
QY 781 atagataagatctcagagctctaataatgctglatcagtaaaaggcaaatcacagat 840
|||||
Db 6003 ATGATTAAGAGCTGAGCTTAAATGTGCTATTCATTAAGGCAAAATCACAGAT 8062
QY 841 ggaagacagattgaaagaaacaaattctcagcctgtaattgatatgttactaaag 900
|||||
Db 6063 GGAAGACGATTGAAAAGGAACAAATTTCATGCAATCTGTAAATTAAGTGTACTAAAAAG 6122
QY 901 aggcagctcaaaagacataccagcttcaacaataaccatactatgaaagagcctc 960
|||||
Db 6123 AGACAGCTCAAGACATATACAGATTAACAAATTAACCTACATATGATGATGATGATGATGAT 6182
QY 961 gaaagaaatcttactagctggtgtcaattgtgaataagtcagagcttttcaagtaatgagaa 1020
|||||
Db 6183 GAAAGAAATCTGAACCATGCTATCTGTAATTAAGATGATGATGATGATGATGATGATGATGAT 6242
QY 1021 gtatctaaagtgatcatatcagctcaacaaagaaacaaatatactcttcttcccaagtg 1080

[illegible]

Db 5284 AACATCTGTTTACAGCTCTAGTACAGAAATTTGGCCAGAAAACCTGACAGCAGA 5343
 Qy 121 atlaagaagaccccttaataatgacatacccttcgaaagaatgttgaagaagctctca 180
 Db 5344 ATTAAGAGCATTTTCATATGCTTCTTCTGAAAAGAAATGCTGAAAGCTTTCTTCAA 5403
 Qy 181 aatgctgataatgacaaaggcagagaatctgtttgttgccttagacagacatcca 240
 Db 5404 AATGCTGATGATGCAAAAGCCAGACAGATGCTGTTGTGTTGATCTAGACAGCATCT 5463
 Qy 241 gttgataagaatatttgatataagttggcccaatgcaaggccagactgtgtgtac 300
 Db 5464 GTTGACCGAATATTGTATGATGATGAGTGGCCCTCTCAAGGGCCACACAGTGTGTTAC 5523
 Qy 301 aacaacacagccatttcaagaagaatgttgaagaatcagaatcttgaagaagcagc 360
 Db 5524 AACACACAGCATTTTACAGAGATGATGTTAGAGAAATTCAGAACTTGGGAAAAGCACC 5583
 Qy 361 aaagaaggaatccctataaactgagacagataggaatagatcaatctgtgtacat 420
 Db 5584 AAGAGAGGAATCTCTGCAAAAACAGACATTTATGAAATGGAATTCATTCCTTTATCAT 5643
 Qy 421 atcaacagactgcccacatcttatttctgtgcaatgacatcctgtgtattttgactcat 480
 Db 5644 ATTACAGACTGCTCTTTTATTTCTGGCAATGACATCTCGGTATTTTGTGATCCCAT 5703
 Qy 481 gccagatatgacacaggggacacatcatcatgagtcggagcagcatgtttagaattggat 540
 Db 5704 GCCAGATATGACACAGAGACCATCATGTTAGCCCTGGACGCGATGTTTAGAGATTTGGAT 5763
 Qy 541 gcaagattttagagacagatctcagatgtctcagatcttattctggaacccaatttaa 600
 Db 5764 GCAGACTTTAGAACCCAGTCTCAGATGTTCTGATCTGACTGTGGAACCACTTTAA 5823
 Qy 601 ctggaataattgcccaatgttcagattctctctcgaatgacaaatgcaagaattcg 660
 Db 5824 CTGGAACAATTTGACATGTTTATGTTCTCTGCGTAATGCAAGATGGCAACAAAGTTTCA 5883
 Qy 661 gaaattctgtcttcacagatcagacagaatgttcagaatcttctggaacaacgagc 720
 Db 5884 GAATTTCTTCCGTTCCATCATCATGACAGAAATGCTCCAGAAATCTTTGGACAAAGTTTACG 5943
 Qy 721 tcaagatggggcagaactcttaattgttcttaatacacaatggaataatcttatttga 780
 Db 5944 TCTGATGGGGCAGAACCTTCTAATGTTTCTCAACACATGAGAAAATATCTATTGTTGAA 6003
 Qy 781 atagaagaagtaactgagctctaaatgtgtgtatctgtaaaaggcaaaatcaaat 840
 Db 6004 ATTAGATTAAGGCCACAGAGAGGTGTAATGTCTTATTCATTAAGGCAAGATCACTGAT 6063
 Qy 841 ggaagacagatggaagaagaacaacttcaatctgttaattgatatgttactataaag 900
 Db 6064 GGAAGACGATTTGAAGAAGCAATTCACAGCCTCTGTAAATTGACAGTGTACTTAAAG 6123
 Qy 901 aggcagctcaaaagacataccagttcaacaataactatactatgtatactgagactct 960
 Db 6124 AGACACTCAAGAGACATACACAGTTCAACAAATACCTACATGATGATCTGAGGATTC 6183
 Qy 961 gaaggaatcttaactcagcgtgactaattgttaataagtcagcttccaagatgagga 1020
 Db 6184 GAAGGAATCTTGACACATAGCTCATCTGTAAATGATGAGATTTTCAAGTATGAAAA 6243
 Qy 1021 gtaactaaagtgtcatcatcagctcaagaagaacagatatctctttcccaagtgtc 1080
 Db 6244 GTATCAAGAGTATATATACAGTCAACAGAACAGATATACCTTTTCCACAGTGT 6303
 Qy 1081 ggaagtgtcgtcgtactcaactaataaanaaaacccatagggtccttctgtt 1140
 Db 6304 GGAGTGTGACAGCTGATTAATCAATATATAAACCACAGAGCCCTTCTCTTCG 6363
 Qy 1141 cctctcttcttgagagctgggtcgcattctcatgtgaatggcactttgacatgattca 1200
 Db 6364 CCTCTCTCTTTTGGAGACAGGGCTGACATTTTCATGTAATGGCCACCTTGTCTAGATTC 6423

Qy 1201 gccagaagacacctgtgcgtgataatgagattgtgttcgaaagtgaactgaataac 1260
 Db 6424 GCCAGAAGAAATCTGTGGCTGATGATTAATGGGGTTGTGTGTAAGTGTGATTAAT 6483
 Qy 1261 agttaatgacagcaataatagctcctgcatatg-tgaattgctaataacagttaaaaa 1319
 Db 6484 AGTTAATGACACATTAATAGCACCTGCATATGTGAGTTACTAATCCAGTTAAAAA 6543
 Qy 1320 cggatttccctgtgtcgtatccacatatagttgttcagagaaccccatcatatgt 1379
 Db 6544 CGGTATTTCCCTGTTGTGACCAACATTAATGATTTTACAGAACACCCATTCAATGTC 6603
 Qy 1380 gtaagagacacttaagaagt 1401
 Db 6604 GTTAAGGACACATTAAGAAGT 6625

RESULT 7
 AB006708
 LOCUS
 DEPOSITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Arabidopsis thaliana (strain: Columbia) DNA, clone: lib.Mitsui P1 clone: MYJ24.
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (sites)
 Kotani, H., Nakamura, Y., Sato, S., Kaneko, T., Asamizu, E., Miyajima, N., and Tabata, S.
 Structural analysis of Arabidopsis thaliana chromosome 5. II. Sequence features of the regions of 1,044,062 bp covered by thirteen physically assigned P1 clones
 DNA Res. 4 (4), 291-300 (1997)
 98069011
 2 (bases 1 to 78844)
 Nakamura, Y.
 Direct Submission
 Submitted (22-AUG-1997) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: ynakam@kazusa.or.jp, Tel: 81-438-52-3935, Fax: 81-438-52-3934)
 Address for correspondence: kaos@kazusa.or.jp
 For the latest information on annotation of this clone, please see <http://www.kazusa.or.jp/kaos/cgi-bin/sgd-graph.cgi?c=MYJ24>
 Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'.
 The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, <http://compbio.ornl.gov/Grail-1.3/>), GENSCAN (Chris Burge, MIT, <http://CCR-081.mit.edu/GENSCAN.html>), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, <http://www.cbs.dtu.dk/services/Netgene2/>) and SplicePredictor (Volker Brendel, Stanford University, <http://grem1ni.zool.iastate.edu/cgi-bin/sp.cgi>).
 Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, <http://genome.wustl.edu/eddy/tRNAscan-SE/>).
 This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions.
 The 5' clone is T2007 and the 3' clone is MKD15.
 Location/Qualifiers
 1..78844
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /chromosome="5"
 /clone="MYJ24"

FEATURES
 source

AGULPETERIGGGGGGDDIMYRAKFERIVGSRDSEAFYMMNPDSNGAPELSIYLLR

Query Match 9.0%; Score 126.6; DB 8; Length 78844;
Best Local Similarity 53.0%; Pred. No. 1,4e-19;
Matches 295; Conservative 0; Mismatches 259; Indels 3; Gaps 1;

Chinese Academy of sciences, 500# Cao Bao Road, Shanghai 200233,
CHINA. E-mail enquiries: bhan@ncgr.ac.cn. Clone requests:
bhan@ncgr.ac.cn

REMARK

Oryza sativa japonica (nipponbare) genomic DNA, chromosome 4, BAC
clone: OSJNB0018M05.

COMMENT

Web site: <http://www.ncgr.ac.cn>
----- Summary Statistics
Assembly program: phrap

* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES

source

1.165299
/organism="Oryza sativa"
/variety="Nipponbare"
/sub_species="japonica"
/db_xref="taxon:4530"
/chromosome="4"
/clone="OSJNB0018M05"
/clone_1ib="CUGI-OSJNBa"

BASE COUNT 46042 a 36580 c 3526 g 47150 t 1 others
ORIGIN

Query Match 8.2%; Score 115.2; DB 2; Length 165299;
Best Local Similarity 52.9%; Pred. No. 5.8e-17;
Matches 271; Conservative 0; Mismatches 238; Indels 3; Gaps 1;

QY 56 atgcaatgctgttttacacacattgacacagaattggcagagaagaataatgacca 115

DB 38846 CAAATGATTTGAGTTTGTCAGGGGTTGCTGAGGCTTTTGACAGCATGAAGATTTCAGTA 38787

QY 116 gcaagaattagagcatccttaatgcatatccttcgaagaagaatgtaagaagcttc 175

DB 38786 CACGACTTAAGCAATCGTTGAGATGATACAGATGACACGAGATTCGTTGAGCTTGG 38727

QY 176 ttcaaatgtctgtatgacgaaggcagagaagaatcgtttgtgttgatcctcagcgc 235

DB 38726 TACAGANTGCAGAAAGATGCTAGAGCTTCTGAGGTTGTGTTTACTGATAGACCCATT 38667

QY 236 atccagttgataagatattgataaagtgtgcccattgcaaggcagcacttgctg 295

DB 38666 ATGCTACTTCACTATTTCTATCACTGGAATGCTGATGAGGAGGCGCTCTCTACT 38607

QY 296 tgtacaacaaccagccattacagaagaatgtagaagaattcagaattcgtgaag 355

DB 38606 GCTTCAATGACTCAATTTTATGCCAGATCTTATGCAATCTCAGCAATTTGCTCAAG 38547

QY 356 gcaagaaagggaagaatccttaaaacgtgacagatagagaatagatcaattcgtgt 415

DB 38546 ACAGCAAACTTGAGAAACCTTTTGCAATAGGACAGGTTGTGCTTGGCTTCAATTGCTAT 38487

QY 416 atcatacacaagatcccatcttatttctgcaagaatgacatcgtgbatlittgac 475

DB 38486 ATCACTTACAGACATTTCCGAGTTGTTCTGG---TGAGATATTTGATATTGTTGANC 38430

QY 476 ctatgcccagatatgacccagggccacatcatttgctccggaagcagatgtaagatt 535

DB 38429 CGATGCTCTTATTCGCGAGAAATATCTCATCTCATGAGGTTCAGAGATAAATTTGG 38370

QY 536 tggatgcagaatttagacacaattcagat 567

DB 38369 TAGGAAGAGAGATCTTAGAACATGTTCCCTGAT 38338

RESULT 9

AC069017 174140 bp DNA linear HMG 12-SEP-2001

LOCUS AC069017 Mus musculus clone MGS3-342116, WORKING DRAFT SEQUENCE, 5 unordered

DEFINITION

ACCESSION AC069017

Accession

Submitted (28-JUL-2000) Han Bin, National Center for Gene Research,

QY 593 attttaaactgataat 609
DB 53423 GTGTGACTTGAGCAT 53439

RESULT 8
OSJNB00002 165299 bp DNA linear HMG 08-SEP-2001

LOCUS Oryza sativa chromosome 4 clone OSJNB0018M05, *** SEQUENCING IN

DEFINITION PROGRESS *** in ordered pieces.

ACCESSION AL606457

VERSION AL606457.1 GI:15552695

KEYWORDS HTG; HTES_PHASE2.

SOURCE Oryza sativa.

ORGANISM Oryza sativa

Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:

Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae:

Embaricoidae: Oryzae: Oryza.

REFERENCE 1 (bases 1 to 165299)

Fu, G., Wang, S. Y., Ren, S. X., Lv, G., Lin, W., Gu, W. Q., Zhu, G. F.,

Tu, Y. F., Jia, J., Yin, H. F., Zhang, Y., Cai, Z., Chen, D., Kang, H.,

Chen, X. Y., Shao, C. Y., Sun, Y., Hu, D. P., Zhang, X. L., Zhang, W.,

Wang, L. J., Ding, C. W., Sheng, H. H., Gu, J. L., Chen, S. T., Ni, L.,

Zhu, F. H., Han, B., Feng, Q., Huang, Y. C., Li, Y., Zhu, J. J., Zhao, Q.,

Hu, X., Liu, Y. L., Mu, J., Yu, Z., Chen, L., Fan, D. L., Zhang, Y. J.,

Zhang, L., Lu, Y. Q., Yu, S. L., Liu, X. H., Lu, T. T., Wang, Y. J., Lu, Y.,

Li, C., Li, T., Zhang, Y., Hu, H., Jia, P. X., Qian, Y. M., Ying, K.,

Zhou, B., Chen, Z. H., Hao, P., Zhang, L., Wu, M., Zhang, R. Q., Guan, J. P.

and Hong, G. F.

Direct Submission

Submitted (28-JUL-2000) Han Bin, National Center for Gene Research,

VERSION AC069017.20 GI:15559167
KEYWORDS HTGS_PHASE1: HTGS_DRAFT: HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus. 1 (bases 1 to 174140)
Metzger, M.L., Lewis, L.R., Hume, J., Edwards, C., Harris, C., Dederich, D., Thomas, S., Okwuonu, G., Carlock, C., Garner, T., Addison, S., Pace, A., Williams, G., Bonnin, D., Brooks, A., Brown, J., Buhay, C., Bunne, C., Burkett, C., Chacko, J., Chen, G., Chen, Z., Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Rocha, S., Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Gill, R., Gorrell, J.H., Gunatane, P., Haller, G., Hernandez, J., Hognes, M., Hosak, H., Hou, X., Huber, J., Jackson, H., Jia, Y., Kelly, J., Kelly, S., Kovak, C., Liu, J., Liu, W., Louised, L., Lozano, R.J., Martin, R., Massery, E., McLeod, M.P., Mei, G., Moore, S., Morgan, M., Morris, S., Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Ogih, M., Parish, B., Perez, L., Reiter, D., Say, J., Shen, H., Vasquez, L., Wallington, S., Williamson, A., Wrensford, G., Zhou, X., Bouck, J., Hodgson, A., Muzny, D.M., Rives, M., Scherer, S., Sodergren, E., Weinstein, G., Worley, K. and Gibbs, R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 174140)
AUTHORS Morley, K.C.

TITLE Direct Submission
JOURNAL Submitted (17-MAY-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Sep 12, 2001 this sequence version replaced gi:14787161.

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information -----
Center project name: MM0
Center clone name: MGS3-342116

----- Summary Statistics -----
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 638 of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 17525 bases at least Q40
Consensus quality: 179254 bases at least Q30
Consensus quality: 181014 bases at least Q20
Estimated insert size: 178322; sum-of-coverage estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 7.4x in Q20 bases; sum-of-coverage estimation

----- NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html). -----
* NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will be preserved.

1 86925: contig of 86925 bp in length
* 86926 87025: gap of unknown length
* 87026 87028: contig of 49703 bp in length
* 136729 136828: gap of unknown length
* 136829 159942: contig of 23114 bp in length
* 159943 160042: gap of unknown length
* 160043 168814: contig of 8772 bp in length
* 168815 168914: gap of unknown length
* 168915 174140: contig of 5226 bp in length.

Location/Qualifiers
1. 174140
/organism="Mus musculus"

FEATURES
source

BASE COUNT 51417 a 32907 c 34252 g 55160 t 404 others
ORIGIN

Query Match 4.1%; Score 58; DB 2; Length 174140;
Best Local Similarity 56.2%; Pred. No. 0.0014; Mismatches 85; Indels 0; Gaps 0;
Matches 109; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 235 catccagatgatagatattgatgatagtgagggcccatgcaaggccagcattgt 294
DB 42084 CATTCGATGACACCCCTGTTAACTAAATGCTTGATTTCAACACCGCTCTAT 42143

QY 295 gtgtacacacacccagcatttaccagaagatgtagtaggaatcgaatcttgaaa 354
DB 42144 TTGTAAACACATGCTGTTTACCCGCCAGAGACTTTCATGCAATGAGAAATGACAA 42203

QY 355 ggcacgaagaagggaatcttataaactggacagatggaatagatcattctgt 414
DB 42204 AGCAGGAAAAAGATGATCCCTGAACTTTGGAAGTTTGAAATGAGTTGCTTC 42263

QY 415 tatcatcacaga 428
DB 42264 TACCATCTAACAG 42277

RESULT 10
PFMAL13P3/c 318221 bp DNA linear HTG 19-AUG-1999
LOCUS Plasmodium falciiparum chromosome 13 strain 3D7, *** SEQUENCING IN PROGRESS ***; in unordered pieces.

ACCESSION AL049184
VERSION AL049184.5 GI:5763803
KEYWORDS HTG: HTGS_PHASE1.
SOURCE malaria parasite P. falciiparum.
ORGANISM Plasmodium falciiparum

REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS Bowman, S., Churcher, C., Harris, B., Harris, D., Lawson, D., Quail, M. and Barrell, B.

TITLE Direct Submission
JOURNAL Submitted (15-MAR-1999) P. falciiparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

COMMENT On Aug 24, 1999 this sequence version replaced gi:5731882.
For more information about this sequence or the Malaria Project, see http://www.sanger.ac.uk/Projects/P_falciiparum. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc.
Order of segments is not known; 800 n's separate segments.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence * as soon as it is available and the accession number will be preserved.

Location/Qualifiers
1. 318221
/organism="Plasmodium falciiparum"
/strain="3D7"
/db_xref="taxon:5833"
/chromosome="13"

BASE COUNT 113610 a 29264 c 27321 g 118405 t 29621 others
ORIGIN

Query Match 3.9%; Score 54.4; DB 2; Length 318221;
Best Local Similarity 49.1%; Pred. No. 0.0092;
Matches 172; Conservative 0; Mismatches 176; Indels 2; Gaps 1;

RESULT 11			
AX083744			
LOCUS	1141 bp	DNA	
DEFINITION	Sequence 22 from Patent WO0111061.	Linear	PAT 28-FEB-2001

VERSION	AX083744.1	GI:13185472
KEYWORDS		
SOURCE		
ORGANISM		
REFERENCE	synthetic construct.	
AUTHORS	synthetic construct	
TITLE	artificial sequence.	
JOURNAL	1 (bases 1 to 1141)	
FEATURES	Kunst, L. and Clemens, S.	
	Regulation of embryonic transcription in plants	
	Patent: WO 011061-A 22 15-FEB-2001;	
	UNIVERSITY OF BRITISH COLUMBIA (CA)	
	Location/Qualifiers	

promoter	/organism="synthetic construct" /db_xref="taxon:32630" 1..1141 /note="consensus sequence of A.t., L.a., and B.n. FAEI Promoters"			
BASE COUNT	123 a	42 g	112 t	832 others
ORIGIN	32 c			

Query Match	3.7%	Score 51.8;	DB 6;	Length 1141;
Best Local Similarity	9.4%	Pred NO. 0.083;		
Matches 75;	Conservative 318;	Mismatches 397;	Indels 6;	Gaps 2;

[illegible]

Db
003 aaagcttcggaaacatcgcgctcgccagcaaccagacagatcgtccagaatccttgaca /12
+ :
131 RAMBDTVDHHYVTAAMNAAWTCMDKDDKRFRWWKKNNNAAGWDGDDTKYHMMNNNCB 190

Dd 191 TWTWVVKYKTDSDSEKRMNYGMBMWKNWSYDVTYYWVWDDMCCKRYKRVWVHTRGRRRN 250

Db 251 YMWAMBTARRRRYNNGWTBAMVYRRMTNNNNNNNNNAKAMCKRKYKVGWNRDABVNSTCTTWK 310

1. **Introduction**
 2. **Background**
 3. **Methodology**
 4. **Results**
 5. **Discussion**
 6. **Conclusion**
 7. **References**
 8. **Appendix**
 9. **Figure 1**
 10. **Figure 2**
 11. **Figure 3**
 12. **Figure 4**
 13. **Figure 5**
 14. **Figure 6**
 15. **Figure 7**
 16. **Figure 8**
 17. **Figure 9**
 18. **Figure 10**
 19. **Figure 11**
 20. **Figure 12**
 21. **Figure 13**
 22. **Figure 14**
 23. **Figure 15**
 24. **Figure 16**
 25. **Figure 17**
 26. **Figure 18**
 27. **Figure 19**
 28. **Figure 20**
 29. **Figure 21**
 30. **Figure 22**
 31. **Figure 23**
 32. **Figure 24**
 33. **Figure 25**
 34. **Figure 26**
 35. **Figure 27**
 36. **Figure 28**
 37. **Figure 29**
 38. **Figure 30**
 39. **Figure 31**
 40. **Figure 32**
 41. **Figure 33**
 42. **Figure 34**
 43. **Figure 35**
 44. **Figure 36**
 45. **Figure 37**
 46. **Figure 38**
 47. **Figure 39**
 48. **Figure 40**
 49. **Figure 41**
 50. **Figure 42**
 51. **Figure 43**
 52. **Figure 44**
 53. **Figure 45**
 54. **Figure 46**
 55. **Figure 47**
 56. **Figure 48**
 57. **Figure 49**
 58. **Figure 50**
 59. **Figure 51**
 60. **Figure 52**
 61. **Figure 53**
 62. **Figure 54**
 63. **Figure 55**
 64. **Figure 56**
 65. **Figure 57**
 66. **Figure 58**
 67. **Figure 59**
 68. **Figure 60**
 69. **Figure 61**
 70. **Figure 62**
 71. **Figure 63**
 72. **Figure 64**
 73. **Figure 65**
 74. **Figure 66**
 75. **Figure 67**
 76. **Figure 68**
 77. **Figure 69**
 78. **Figure 70**
 79. **Figure 71**
 80. **Figure 72**
 81. **Figure 73**
 82. **Figure 74**
 83. **Figure 75**
 84. **Figure 76**
 85. **Figure 77**
 86. **Figure 78**
 87. **Figure 79**
 88. **Figure 80**
 89. **Figure 81**
 90. **Figure 82**
 91. **Figure 83**
 92. **Figure 84**
 93. **Figure 85**
 94. **Figure 86**
 95. **Figure 87**
 96. **Figure 88**
 97. **Figure 89**
 98. **Figure 90**
 99. **Figure 91**
 100. **Figure 92**
 101. **Figure 93**
 102. **Figure 94**
 103. **Figure 95**
 104. **Figure 96**
 105. **Figure 97**
 106. **Figure 98**
 107. **Figure 99**
 108. **Figure 100**
 109. **Figure 101**
 110. **Figure 102**
 111. **Figure 103**
 112. **Figure 104**
 113. **Figure 105**
 114. **Figure 106**
 115. **Figure 107**
 116. **Figure 108**
 117. **Figure 109**
 118. **Figure 110**
 119. **Figure 111**
 120. **Figure 112**
 121. **Figure 113**
 122. **Figure 114**
 123. **Figure 115**
 124. **Figure 116**
 125. **Figure 117**
 126. **Figure 118**
 127. **Figure 119**
 128. **Figure 120**
 129. **Figure 121**
 130. **Figure 122**
 131. **Figure 123**
 132. **Figure 124**
 133. **Figure 125**
 134. **Figure 126**
 135. **Figure 127**
 136. **Figure 128**
 137. **Figure 129**
 138. **Figure 130**
 139. **Figure 131**
 140. **Figure 132**
 141. **Figure 133**
 142. **Figure 134**
 143. **Figure 135**
 144. **Figure 136**
 145. **Figure 137**
 146. **Figure 138**
 147. **Figure 139**
 148. **Figure 140**
 149. **Figure 141**
 150. **Figure 142**
 151. **Figure 143**
 152. **Figure 144**
 153. **Figure 145**
 154. **Figure 146**
 155. **Figure 147**
 156. **Figure 148**
 157. **Figure 149**
 158. **Figure 150**
 159. **Figure 151**
 160. **Figure 152**
 161. **Figure 153**
 162. **Figure 154**
 163. **Figure 155**
 164. **Figure 156**
 165. **Figure 157**
 166. **Figure 158**
 167. **Figure 159**
 168. **Figure 160**
 169. **Figure 161**
 170. **Figure 162**
 171. **Figure 163**
 172. **Figure 164**
 173. **Figure 165**
 174. **Figure 166**
 175. **Figure 167**
 176. **Figure 168**
 177. **Figure 169**
 178. **Figure 170**
 179. **Figure 171**
 180. **Figure 172**
 181. **Figure 173**
 182. **Figure 174**
 183. **Figure 175**
 184. **Figure 176**
 185. **Figure 177**
 186. **Figure 178**
 187. **Figure 179**
 188. **Figure 180**
 189. **Figure 181**
 190. **Figure 182**
 191. **Figure 183**
 192. **Figure 184**
 193. **Figure 185**
 194. **Figure 186**
 195. **Figure 187**
 196. **Figure 188**
 197. **Figure 189**
 198. **Figure 190**
 199. **Figure 191**
 200. **Figure 192**
 201. **Figure 193**
 202. **Figure 194**
 203. **Figure 195**
 204. **Figure 196**
 205. **Figure 197**
 206. **Figure 198**
 207. **Figure 199**
 208. **Figure 200**
 209. **Figure 201**
 210. **Figure 202**
 211. **Figure 203**
 212. **Figure 204**
 213. **Figure 205**
 214. **Figure 206**
 215. **Figure 207**
 216. **Figure 208**
 217. **Figure 209**

RESULT	12
AF151112	
LOCUS	2389 bp mRNA linear INV 29-MAR-2000
DEFINITION	Dicystosellium discoideum cheater Chla (cht) mRNA, cht-A allele,
	partial cds.
ACCESSION	AF151112
VERSION	AF151112.1 GI:5007063

SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE
Dictyostelium discoideum.	Dictyostelium discoideum.	Dictyostellida: Dictyostellium.	Emmis,H.L., Dao,D.N., Pukatzki,S.U. and Kessin,R.H.	Dictyostellium amoebae lacking an F-box protein form spores rather than stalk in chimerae with wild type	Proc. Acad. Sci. U.S.A.	97 (7), 3292-3297 (2000)
20202628						

AUTHORS	Dao, D., Emmls, H. L., Franke, J., Kessin, R. H., Nelson, M. K. and Pukatzki, S.
TITLE	A F-box protein is essential for development in Dictyostellium discoideum
JOURNAL REFERENCE	Unpublished (bases 1 to 2389)
AUTHORS	Dao, D., Emmls, H. L., Franke, J., Kessin, R. H., Nelson, M. K. and Pukatzki, S.
TITLE	Direct Submission
JOURNAL	Submitted (13-MAY-1999) Anatomy and Cell Biology, Columbia University, 630 W 168 Street, New York, NY 10032, USA
FEATURES	Location/Qualifiers

```
/organism="Dictyostelium discoideum"  
/strain="AX3K"  
/db_xref="taxon:4689"
```


	gene	/chromosome="2"
	CDS	1..>2389 /gene="cht" /allele="A" 230..>2389 /gene="cht" /note="F-box protein with WD40 domains" /codon_start=1 /product="cheater ChtA" /protein_id="AAD37800.1" /db_xref="GI:5007064"
	BASE COUNT	990 a 305 c 369 g 725 t
	ORIGIN	LHFHKTPTKRQIIHSNLSDENKYNVNSGSDAGSSSQ"
	Query Match	3.2% Score 45; DB 3; Length 2389;
	Best Local Similarity	49.0%; Pred. No. 2.9;
	Matches 120; Conservative	0; Mismatches 125; Indels 0; Gaps 0;
QY	694 gtccagaatccttcttgacaactgcgcccagatggaggaggaactctaagtcttaatt 753	
Dd	1583 GTTCAACATCTATTGGTGGAATGTGGTTCTCATATATGTAATCCTCCTCAAGTCATATAATGCCA 1642	
QY	754 cacatygaaaaaatcttattcttcttgataagataagaaactcgaatcgctgaatgctg 813	
Dd	1643 ATGGCGAAATTTTGGTGAATATACTTTTAGATGTAAGTAAATGAAGTAAAGTAAAGTAAAGTAAAGTAA 1702	
QY	814 tattcaagtaaaggacaaaatacacagatgtagagacgatgtaagaagaaagaaatctcatgca 873	
Dd	1703 TCTTCAAAAATAATTAAGAATAAAGAAATTAATAATATGAATAATTAATAATATATGAT 1762	
QY	874 tccttaatttgtatgtttactaataaaggagagctccaagaacataccagcttcatacaata 933	
Dd	1763 AATATATATTGATGAAGAATTTGAACAAGAGGAGAAAAATATTGAAAAAGATCAACACAA 1822	
QY	934 accta 938	
Dd	1823 ACCCA 1827	
RESULT 13	AFI51733	
LOCUS	AFI51733	3885 bp DNA linear INV 07-MAY-2001
DEFINITION	Dictyostelium discoidelium F-Box A protein (fbxa) gene, complete cds.	
ACCESSION	AFI51733	
VERSION	AFI51733.1	GI:5163417
KEYWORDS		
SOURCE	Dictyostelium discoideum.	
ORGANISM	Dictyostelium discoideum.	
AUTHORS	Nelson,M.K., Clark,A., Abe,T., Nomura,A., Yadava,N., Funai,I,C.J., Jermyu,K.A., Mohanty,S., Fitrel,R.A. and Williams,J.G. An F-box/WD40 repeat-containing protein important for Dictyostelium cell-type proportioning, slug behaviour, and culmination Dev. Biol. 224 (1), 42-59 (2000)	
JOURNAL	20359235	
MEDLINE	PUBMED	
REFERENCE	2 (bases 1 to 3885) Nelson,M.K., Clark,A., Abe,T., Nomura,A., Fitrel,R.A. and	

[illegible]


```
/clone="RP3-425C14"
/clone_lib="RPCT-3"
complement(1..61)
/Note="AluYb8 repeat: matches 62..3 of consensus"
535..837
/Note="AluYb repeat: matches 1..302 of consensus"
complement(1572..2210)
/Note="match: GSS A0013878"
2351..2483
/Note="FLAM_A repeat: matches 1..133 of consensus"
complement(4425..4846)
/Note="LIME3A repeat: matches 457..5 of consensus"
complement(4987..5365)
/Note="MLT1A2 repeat: matches 374..9 of consensus"
complement(7622..8574)
/Note="LIMB5 repeat: matches 921..5 of consensus"
complement(8424..9042)
/Note="LI repeat: matches 5384..4729 of consensus"
complement(9285..9672)
/Note="match: GSS A0015116"
join(<9584..9692,9848..9975,10736..10860,13432..13507,
16379..16440,17346..17433,19361..19509,19929..20168,
20792..20897,25114..25167,28641..28725,29136..30177)
/Note="match: CDNA5 M65217 X61754 I06125: match: ESTs
AA832774 AA913486 AA913304 H72024 AA832045 R46149 R21303
AA868883 D52415 D52616 AA944812 C20787 H14792 N39221
D53854 AA551748 AA551757 D52419 R14019 T53313 AA250770
AA651949 AA969316 AA707596 AA860985 R39144 AA253434 H24055
AA693992 AA815317 N46665 T29491 AA905319 AA294840 H14791
H24163 AA136430 AA329693 W22062 AA334787 AA365830"
/product="dJ425C14.1 (Heat Shock Factor 2 (HSF2, Heat
Shock Transcription Factor 2, HSF 2))"
/evidence=not_experimental
9584..30177
/Note="HSF2"
join(<9584..9692,9848..9975,10736..10860,13432..13507,
16379..16440,17346..17433,19361..19509,19929..20168,
20792..20897,25114..25167,28641..28725,29136..29431)
/Note="HSF2"
/Note="match: proteins P38533 Q00393 P38530 P41154 Q00613
P38529 P38531 Q63717 P38532 Q99472"
/codon_start=1
/evidence=not_experimental
/product="dJ425C14.1 (Heat Shock Factor 2 (HSF2, Heat
Shock Transcription Factor 2, HSF 2))"
/protein_id="CAB16203.1"
/db_xref="GI:3294179"
/translation="NGOSFLVLDGEORFAKEILLPKYKHNMA5FVROLNMYGRKYVH
INDGIYKQERDGPVERQHPYFKOGDDLEINIRKYS5SKPRENKIRQEDLITRIS5A
QKVOIKQETESRLSKSENE5LMEVESELRAKHAQOQVIRKIVQFTVTLVONNOL
VSLKRRRLPLNTNGAOKNLFQHIKVEPTDNHHKVP5RTEGLKPREISIDIIY
DYTDNADADENIPVETNEEDVSDPNSQYPDIVIEDNEDAPVQ5QSEONEP
AAESLSSGDSGSP5LMSAVOLN5SSLSSEDPVMDSTINDINILKVELLDVLD
SIDSLIEDFOAM5GROF5IPDLVILFT5SVQMPYQYINTK5ENK5E5ETTKNV
VQPV5EGRK5K5K5PKO5IOTVTA5PLFLA5DNP5P55QOAT5T5AS5EVL5VD5PI
EYDELID5SLD5EP5Q5K5L5R5E5PL5TEA5R5E5T5L5E5L5A5P5LD5D5P5L5D5"
complement(11178..11670)
/Note="match: GSS B51948"
complement(11840..12127)
/Note="AluX repeat: matches 300..1 of consensus"
12541..12723
/Note="MIR repeat: matches 71..251 of consensus"
12944..13243
/Note="AluX repeat: matches 2..292 of consensus"
complement(14421..14487)
/Note="MER5B repeat: matches 74..7 of consensus"
complement(15495..15671)
/Note="MER5A repeat: matches 189..2 of consensus"
complement(15855..15982)
/Note="FLAM_C repeat: matches 128..1 of consensus"
/Note="match(17921..18221)
"tches 302 2 of consensus"

repeat_region 22624..22769
/Note="MIR repeat: matches 49..199 of consensus"
complement(26830..27536)
/Note="LI repeat: matches 4031..3265 of consensus"
29432..29832
/Note="HSF2"
/Note="match: STS G13562"
complement(29857..30169)
/Note="match: STS G05769"
30148..30153
/Note="HSF2"
30171
/Note="HSF2"
30177
/Note="HSF2"
complement(31226..31515)
/Note="AluX repeat: matches 292..2 of consensus"
complement(32572..32871)
/Note="AluX repeat: matches 301..1 of consensus"
32902..33204
/Note="AluY repeat: matches 1..299 of consensus"
/Note="AluY repeat: matches 1..299 of consensus"
34025..34261
/Note="match: STS G24632"
34383..34412
/Note="15 copies 2 mer tg 93% conserved"
35013..35126
/Note="LI repeat: matches 2394..2506 of consensus"
35128..35538
/Note="MSTRA repeat: matches 1..426 of consensus"
35539..35586
/Note="MSTRA repeat: matches 1..48 of consensus"
35578..36192
/Note="LI repeat: matches 1057..1737 of consensus"
35600..36060
/Note="MER25 repeat: matches 1583..2099 of consensus"
36219..37029
/Note="MST-INTERNAL repeat: matches 807..1651 of
consensus"
37030..37447
/Note="MSTB repeat: matches 1..426 of consensus"
37541..38260
/Note="LI repeat: matches 3171..3958 of consensus"
38261..38558
/Note="AluX repeat: matches 1..302 of consensus"
38566..39641
/Note="LI repeat: matches 3943..5041 of consensus"
complement(39724..40048)
/Note="LIMD1 repeat: matches 749..390 of consensus"
complement(join(40559..42390,43985..44215,44332..44476,
48875..48965,49099..49268,50981..51118,51362..51441,
53692..53861,55731..555892))
/Note="match: CDNA5 U49188 L29441: match: ESTs R63311
AA642883 AA680098 AA757435 AA147658 AA364378 N22238 R73280
AA563775 AA479364 N68017 H96625 R64053 R62561 AA054470
N39921 AA587827 AA718993 AA909130 D55889 AA802280 R80437
AA661998 AA056751 R63339 AA946958 AA788657 N47022 AA855136
AA640774 AA359308 T77723 N24641 D60199 C14222 AA393746
AA649582 AA838780 W03704 AA055102 T30549 N75278 AA903139
R74572 C16158 AA573354 C16394 AA658814 AA706277 R34117
AA582675 H34768 A1003241 T34540 T10566 W03066 M39974
AA425614 W32165 C18045 W23860 AA792501 N31789 N87445
AA384285 AA238110 R10545 R57569 AA099167 H96624 H47943
W76583 AA298096 AA825649 Z42747 T32615 H47954 N31858
H70513 AA010230 W02277 AA178232 AA298204 AA462364 AA178351
R62199 AA318861 Z44998 D60904 D52653 R12298 AA137672
R11687 C82588 H35640 AA590123 W97491 AA633788 R18882
AA658854 AA657717 AA657439 AA000956 AA666318 AA087604
W52428 T59284 AA068123 AA517373 T59243 AA912593 AA836161
AA984271 AA982116 AA096521 AA852720 AA986636 AA853329
AA344622"
/product="dJ425C14.2 (Placental protein DIFF33 LINE)"
/evidence=not_experimental
```

559. .55892)

④

Query Match

Query Match	3.2%	Score 44.6;	DB 9;	Length 160203;
Best Local Similarity	44.2%;	Pred. NO. 2;		
Matches 185; Conservative	0;	Mismatches 234;	Indels 0;	Gaps 0;

OY	136	aatgcattcccttcctgaagaagaaagtgttgaagaagcctctccaatgctgatgatgca	195
Db	5924	AAGCATTAAAGGCTGTGAAGAGAAGAAAATTAATGTTCATCCACGACTACTATGGGTGTTCT	5865
OY	196	aaggcgacagaatacttglttttgttgaacctagacagcatccacttgatagaatatlt	255
Db	5864	CAGAAACTGAATCAAGTAAATGATTTCCTCACGTGATTTATTATCATATATAAAAAATT	5805
OY	256	gatgataagtgggcccatctgcaaggcgcaagcatttgtgttacacaaccagccatt	315
Db	5804	ACTTGTTTTCCTAATTACGCACAACAGCTTAGAATCTATTATTTAAGTCAGAACACCATT	5745
OY	316	acagaagatglttgtaggaatctcgaatcttgaaaaagcagaagaagggaatatct	375
Db	5744	TACAACTATATATAGATATCCAGAGAAAGAAATGAACAAAGATCCACAGGCCCTAATTAGCA	5685
OY	376	tataaaactgagcaatgtaatgaatcataactctgttatcatatcacagactgccca	435
Db	5684	AACATTAATCACAAATGSAATTTAAATAATATGGCCCTTGTCATGAGAAATACMAACACAGCA	5625
OY	436	tctttatcttcggcaatgacatccctgtatltttgatccatcgacagabatyacca	495
Db	5624	GATATCATTTTTCCCTAATCTCGATTTATATAGTTTGTGATCCAAATGCAGCAAAAATCCAA	5565
OY	496	ggggcacatcatcatlgtcccgcagcagatgttagagatttgatgcagatalltagaac	554
Db	5564	AAGCCTTTTTTTTTTATATGAACTTGATTTAAACATATGTGGAAGAGCTGTAAGGC	5506

Search completed: May 22, 2002, 06:53:41
Job time: 8806 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 06:43:32 ; Search time 463.88 Seconds
(without alignments)
5185.387 Million cell updates/sec

Title: US-09-693-205-7_COPY_5300_6700

Sequence: 1 gtacgcagtaaacatagagc.....aaagacacttaagaagt 1401

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_032802.*
1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1401	100.0	12792	22	AAH20176	Human mutated spas
2	1390	99.2	12793	22	AAH20174	Human spastin nucl
3	1390	99.2	12793	22	AAH20178	Human mutated spas
4	1390	99.2	12793	22	AAH20182	Human mutated spas
5	1388.4	99.1	12793	22	AAH20179	Human mutated spas
6	1174	83.8	11493	22	AAH20175	Human spastin nucl
7	40.6	2.9	7990	24	ABJL32158	Mouse spastin
8	39.6	2.8	531	22	ABA59780	Human immune syste
9	39.6	2.8	531	22	ABA28269	Human foetal liver

C	10	39.6	2.8	531	22	AAK08051	Human brain expres
C	11	39.6	2.8	531	22	AAK33925	Human bone marrow
C	12	39.6	2.8	531	22	AAI16573	Probe #6506 for ge
C	13	39.6	2.8	531	22	AAI39650	Probe #8336 used t
C	14	39.6	2.8	10119	22	AAK87559	Human immune/haema
C	15	39.6	2.8	12247	22	AAK73076	Human immune/haema
C	16	39.6	2.8	12247	22	AAK87543	Human immune/haema
C	17	39.6	2.8	14983	22	AAK73075	Human immune/haema
C	18	39.6	2.8	14983	22	AAK87542	Human immune/haema
C	19	39.6	2.8	20991	22	AAK87545	Human immune/haema
C	20	39.6	2.8	32190	22	AAI62927	Human immune/haema
C	21	39.6	2.8	32249	22	AAI62932	Human genomic DNA
C	22	39.6	2.8	39068	22	AAK71820	Human immune/haema
C	23	39.6	2.8	39068	22	AAK73078	Human immune/haema
C	24	39.6	2.8	39068	22	AAK85294	Human immune/haema
C	25	39.6	2.8	39068	22	AAK87544	Human immune/haema
C	26	39.6	2.8	39110	22	AAK71825	Human immune/haema
C	27	39.6	2.8	39110	22	AAK73087	Human immune/haema
C	28	39.6	2.8	39110	22	AAK87555	Human immune/haema
C	29	39.6	2.8	45300	22	AAK73079	Human immune/haema
C	30	39.6	2.8	45300	22	AAK87547	Human immune/haema
C	31	39.4	2.8	815	21	AAF21681	Human breast and o
C	32	39	2.8	7503	21	AAAY0206	Human Plasmodium falci
C	33	38.6	2.8	1092	23	AA555809	Human Streptococcus pneu
C	34	38	2.7	21358	22	AA539919	Human genomic sequenc
C	35	38	2.7	21358	22	AAK73090	Human reproductive
C	36	38	2.7	21358	22	AAK73090	Human reproductive
C	37	38	2.7	21358	22	AAK87446	Human immune/haema
C	38	38	2.7	21358	22	AAK87558	Human immune/haema
C	39	38	2.7	21358	22	AAK90363	Human immune/haema
C	40	38	2.7	21676	22	AA539918	Human digestive sy
C	41	38	2.7	21676	22	AAI06418	Human reproductive
C	42	38	2.7	21676	22	AAK73081	Human immune/haema
C	43	38	2.7	21676	22	AAK87445	Human immune/haema
C	44	38	2.7	21676	22	AAK87549	Human immune/haema
C	45	38	2.7	21676	22	AAK90362	Human digestive sy

ALIGNMENTS

RESULT 1	
ID	AAH20176 standard; DNA: 12792 BP.
AAH20176	
AC	AAH20176;
XX	
DT	09-AUG-2001 (first entry)
XX	
DE	Human mutated spastin nucleotide sequence SEQ ID NO:7.
XX	
KW	Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
KW	autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KW	neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KW	reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
KW	atrophy of upper cerebellar vermis; absence of Purkinje cell;
KW	abnormal neuronal lipid storage; genetic disorder; characterisation: ds.
XX	
OS	Homo sapiens.
OS	Synthetic.
XX	
FT	Key
FT	CDS
FT	Location/Qualifiers
FT	/*tag= a
FT	/product= "mutated spastin"
MO200129266-A2.	
26-APR-2001.	
20-OCT-2000; 2000WO-US29130.	
20-OCT-1999; 99US-0160588.	

XX (UYMC-) UNIV MCGILL.
PA (HOPI-) HOPITAL SAINTE-JUSTINE.
XX
XX
PI Hudson TJ, Engert J, Richter A;
XX
DR WPI: 2001-308494/32.
DR P-PSDB; AAB97821.
XX
XX
PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -
XX
PS Claim 1; Page -: 76pp; English.

CC The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSAACS)) gene sequences (1). The spastin gene has been mapped to
CC chromosome 13q11. (1) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (1), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (1) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (1) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (1). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with Spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes a mutated human spastin.
CC N.B. The present sequence is not given in the present specification but
CC is derived from the human spastin nucleotide sequence (AAH20174) as
CC stated on page 14.

Sequence 12792 BP; 4163 A; 2256 C; 2487 G; 3886 T; 0 other;

Query Match	100.0%;	Score 1401;	DB 22;	Length 12792;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1401; Conservative	0;	Mismatches	0;	Gaps 0;

QY	1	gttcgcagtaaacttagtagagcagctcccaagcgacacacaagaaccttagaagaatatgcatacc	60
Db	5300	gttagcagataaactaagtagcagctcccaagcgacacacaagaaccttagaagaatatgcatacc	5359
QY	61	aatgtcgtgtttaacaaaccttggcacagaaatttgggcacagaagaataattgacacagaga	120
Db	5360	aatgtcgtgttttaacaaaccttggcacagaaatttgggcacagaagaataattgacacagaga	5419
QY	121	attaagagcatcccttaataatcatctcttcgaaagaagaaatgttgaagaagctcttccaa	180
Db	5420	attaagagcatcccttaataatcatctcttcgaaagaagaaatgttgaagaagctcttccaa	5479
QY	181	aatgtcgtatgatgcagaagcgacagaaatctgttttgggttgatccataagacatcca	240
Db	5480	aatgtcgtatgatgcagaagcgacagaaatctgttttgggttgatccataagacatcca	5539
QY	241	gtttgatagaatatcttgaatgaatgaatggtggcccatctgcagaagccagcaacttgtgttac	300
Db	5540	gtttgatagaatatcttgaatgaatgaatggtggcccatctgcagaagccagcaacttgtgttac	5599
QY	301	aacaaacacgacattatcacagaagatgtgttaagggaaatcaagaatcttggaaaaagcagc	360
Db	5600	aacaaacacgacattatcacagaagatgtgttaagggaaatcaagaatcttggaaaaagcagc	5659

QY	361	aaagagggnaaactccataaaactggacagatgaaagatgaaatccaatctggtatcat	420
Db	5660	aaagagggnaaactccataaaactggacagatgaaagatgaaatccaatctggtatcat	5719
QY	421	atccagacatgccacattttattcttcggaaatgaaatccctgatatcttggatcccat	480
Db	5720	atccagacatgccacattttattcttcggaaatgaaatccctgatatcttggatcccat	5779
QY	481	gccagatatgacacaggggccacacatcatgctcccgagcgagttctaaagatttggat	540
Db	5780	gccagatatgacacaggggccacacatcatgctcccgagcgagttctaaagatttggat	5839
QY	541	gccatttttggaaacagcttctaagttcttggatcttattcttgggaacccatttaa	600
Db	5840	gccatttttggaaacagcttctaagttcttggatcttattcttgggaacccatttaa	5889
QY	601	ctggataattgcaacaagtttcagatcttcctcttcgtaatgcgaagaaatggcaaaatttcg	660
Db	5900	ctggataattgcaacaagtttcagatcttcctcttcgtaatgcgaagaaatggcaaaatttcg	5959
QY	661	gaaatttcgctctgtccagacatcaagacgaatggtccagaaatcttttggacaactgggc	720
Db	5960	gaaatttcgctctgtccagacatcaagacgaatggtccagaaatcttttggacaactgggc	6019
QY	721	tcaagttggggcgaacacttccaatgtttcttaatcacatgaaanaaatcttcattgtgaa	780
Db	6020	tcaagttggggcgaacacttccaatgtttcttaatcacatgaaanaaatcttcattgtgaa	6079
QY	781	atagataagagttactgtagctctaaatgtgctgatatctgaataaggccaataatccagat	840
Db	6080	atagataagagttactgtagctctaaatgtgctgatatctgaataaggccaataatccagat	6139
QY	841	ggagacagattgtaaaaggaaacacattctcatgcactcgttaattgaaagttaaccaaag	900
Db	6140	ggagacagattgtaaaaggaaacacattctcatgcactcgttaattgaaagttaaccaaag	6139
QY	901	aggcagctcctcaaaaggacatccagttccaacaaataacctatactatgatacttgaagactct	960
Db	6200	aggcagctcctcaaaaggacatccagttccaacaaataacctatactatgatacttgaagactct	6259
QY	961	gaaggaatcttactaagcttgagcgaattttgtaataagatcagagcttccaagatgagaa	1020
Db	6260	gaaggaatcttactaagcttgagcgaattttgtaataagatcagagcttccaagatgagaa	6319
QY	1021	gtactctaaagtgtcatatcagcttcacaaagacccaagatatctactctttcccagtggt	1080
Db	6320	gtactctaaagtgtcatatcagcttcacaaagacccaagatatctactctttcccagtggt	6379
QY	1081	ggagtaagctgcctgcatattactccaactataaaaaaaccocataggccttctgttttttg	1140
Db	6380	ggagtaagctgcctgcatattactccaactataaaaaaaccocataggccttctgttttttg	6439
QY	1141	ccctcttcttggagacttggcgctgcacatttcagttaaatgagcgaacttgcgaactgatatca	1200
Db	6440	ccctcttcttggagacttggcgctgcacatttcagttaaatgagcgaacttgcgaactgatatca	6499
QY	1201	ggcgaaggaacactgtgcggtgaatgaatggagttgtgttcgaagtgcacttggataaac	1260
Db	6500	ggcgaaggaacactgtgcggtgaatgaatggagttgtgttcgaagtgcacttggataaac	6559
QY	1261	agtttaatgacagatttaaatagctccctgcgataatgtaattgttaataaagttaaaaaac	1320
Db	6560	agtttaatgacagatttaaatagctccctgcgataatgtaattgttaataaagttaaaaaac	6619
QY	1321	ggtatcttccctggtctgataccaactatcaagtgttaagaagaccccatcttatgttg	1380
Db	6620	ggtatcttccctggtctgataccaactatcaagtgtgttaagaagaccccatcttatgttg	6679
QY	1381	taaaagcactttaaagaagt 1401	
Db	6680	taaaagcactttaaagaagt 6700	

Oy	1381	t	a	a	g	g	a	c	a	c	t	t	a	a	g	a	a	g	t	1401
Db	6680	t	a	a	g	a	c	a	c	t	t	t	a	a	g	a	a	g	t	6700

Result	2	
AAH20174	AAH20174	standard; DNA: 12793 BP.
ID	AAH20174	standard; DNA: 12793 BP.
XX	AAH20174;	
AC	AAH20174;	
XX	09-AUG-2001	(first entry)
DT	09-AUG-2001	(first entry)
XX	Human spastin nucleotide sequence SEQ ID NO:1.	
DE	Human: mouse; spastin; ARSACS; chromosome 13q11; identification;	
XX	autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;	
KW	neurodegenerative disease; reduced sensory nerve conduction; diagnosis;	
KW	reduced motor nerve velocity; hypermyelination of retinal nerve fibre;	
KW	atrophy of upper cerebellar vermis; absence of Purkinje cell;	
KW	abnormal neuronal lipid storage; genetic disorder; characterisation; ds.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	77..11566
FT		/*tag= a
FT		/product= "spastin"
XX		
PN	MO200129266-A2.	
XX		
PD	26-APR-2001.	
XX		
PF	20-OCT-2000; 2000MO-US29130.	
XX		
PR	20-OCT-1999; 99US-0160588.	
XX		
PA	(UYMC-) UNIV MCGILL.	
PA	(HOP1-) HOPITAL SAINTE-JUSTINE.	
XX		
PI	Hudson TJ, Engert J, Richter A;	
DR	WPI: 2001-308496/32.	
XX	P-PSDB: AAB97819.	
XX		
PT	New isolated polynucleotide, encoding spastin gene, and polypeptides,	
PT	useful for diagnosing autosomal recessive spastic ataxia of	
PT	Charlevoix-Saguenay disease by detecting two point mutations in spastin	
PT	gene sequence -	
XX		
PS	Claim 1; Fig 9; 76pp; English.	
XX		
CC	The present invention describes human and mouse spastin, and mutated	
CC	human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay	
CC	(ARSACS) gene sequences (I). The spastin gene has been mapped to	
CC	chromosome 13q11. (I) have neuroprotective activities and can be used in	
CC	gene therapy and as a spastin polypeptide agonists. (I), their fragments	
CC	or their complements can be useful for assaying the presence of a nucleic	
CC	acid molecule in a sample. (I) is useful for diagnosing or aiding in the	
CC	diagnosis of an early onset neurodegenerative disease in an individual.	
CC	The neurodegenerative disease comprises reduced sensory nerve conduction	
CC	reduced motor nerve velocity, hypermyelination of retinal nerve fibres,	
CC	atrophy of upper cerebellar vermis, absence of Purkinje cells and	
CC	abnormal neuronal lipid storage. (I) can also be used to produce	
CC	antisense nucleic acids, is useful as molecular weight or chromosome	
CC	markers, to identify genetic disorders, as hybridisation probes or	
CC	primers, as an antigen, identify and express recombinant protein for	
CC	analysis, characterisation or therapeutic use, or as markers for tissues	
CC	in which the corresponding protein is expressed. Diagnostic methods from	
CC	the present invention can be used to identify subjects having or at risk	
CC	of developing a disease or disorder associated with aberrant expression	
CC	or activity of (I). The assays can be utilised to identify a subject	
CC	having or at risk of developing a disorder associated with spastin	
CC	protein or spastin gene expression or activity. The present sequence	
CC	encodes human spastin as given in the present invention.	
XX		
Sequence	12793 BP; 4163 A; 2256 C; 2487 G; 3887 T; 0 other;	

Query Match		99.2%	Score 1390	DB 22	Length 12793
Best Local Similarity		99.9%	Pred. No. 0		
Matches 1401		Conservative	0	Mismatches	0
				Indels	1
				Gaps	
OY	1	gtacacgtaaacatcagagcagtcgcccaaaacgcacaaaagccttagaagatgcatcc	60		
Db	5300	gtacagttaaacactagagacagtcgcccaaaacgcacaaaagccttagaagatgcatcc	53559		
OY	61	aagtcgtgttttacaaacacttggcacagaatttggcagaagaattgacagcaga	120		
Db	5360	aatgtcgttttacaaacacttggcacagaatttggcagaagaattgacagcaga	5419		
OY	121	atthaagacatcccttaatgtatatactcttcgaaagaagaatttgaagagctttcaa	180		
Db	5420	atthaagacatcccttaatgtatatactcttcgaaagaagaatttgaagagctttcaa	5479		
OY	181	aatgtcatgatgcagaagcgacagaaatctgttttggtttctgactcttagaacatcca	240		
Db	5480	aatgtcatgatgcagaagcgacagaaatctgttttggtttctgactcttagaacatcca	5539		
OY	241	gtctatgataatattgtgtataagvtggcccatctgcaggccagcacttgtgtac	300		
Db	5540	gtctatgataatattgtgtataagvtggcccatctgcaggccagcacttgtgtac	5599		
OY	301	aacacccagccatttacagaagaatgttttagaggaattcagaactcttgaaaaagcag	360		
Db	5600	aacacacccagccatttacagaagaatgttttagaggaattcagaactcttgaaaaagcag	56559		
OY	361	aaaagagggaaatcccttaaaacttgcagatgcatagcataatctgtgtatcat	420		
Db	5660	aaaagagggaaatcccttaaaacttgcagatgcatagcataatctgtgtatcat	5719		
OY	421	atccagagatggcccatctttattcttgcgaatgcatactctgtatatttgatcccat	480		
Db	5720	atccagagatggcccatctttattcttgcgaatgcatactctgtatatttgatcccat	5779		
OY	481	ggcagatatgcacaaaggggcacacatcatagtcgcccgacgcagtltttagaatttgat	540		
Db	5780	ggcagatatgcacaaaggggcacacatcatagtcgcccgacgcagtltttagaatttgat	5839		
OY	541	gcagattttagacacagttctcgaatgttctgatatcttatactcttgcgaaacccatttaa	600		
Db	5840	gcagattttagacacagttctcgaatgttctgatatcttatactcttgcgaaacccatttaa	5899		
OY	601	ctgagataatgcacaagttccagatctccctctgcgaatgcagaagaatgycaaaagtctcg	660		
Db	5900	ctgagataatgcacaagttccagatctccctctgcgaatgcagaagaatgycaaaagtctcg	59559		
OY	661	gaaattcgtctgttccagatcatcagaagaatggtccagaatcttcttgacaaactgcgc	720		
Db	5960	gaaattcgtctgttccagatcatcagaagaatggtccagaatcttcttgacaaactgcgc	6019		
OY	721	tcaagatggggagagacacttctaaagtgttcttatacacaatggaaaaattcttatgttga	780		
Db	6020	tcaagatggggagagacacttctaaagtgttcttatacacaatggaaaaattcttatgttga	6079		
OY	781	ataataagagttactgtgagctctaaatgtgtcttatacttcagtaaagggcacaatacagat	840		
Db	6080	ataataagagttactgtgagctctaaatgtgtcttatacttcagtaaagggcacaatacagat	6139		
OY	841	ggagacacagatctgaaaaggaaaacttcataatgcatactctgtataatgtatcaggttaaaag	900		
Db	6140	ggagacacagatctgaaaaggaaaacttcataatgcatactctgtataatgtatcaggttaaaag	6199		
OY	901	aggcagctcacaagacataccagttccaacaataactataactatactatgatactgagacact	960		
Db	6200	aggcagctcacaagacataccagttccaacaataactataactatactatgatactgagacact	62559		
OY	961	gaaggaatctctactacgtgtgctaattctgtaatagatcagccttccaagtatggagaa	10200		
Db	6260	gaaggaatctctactacgtgtgctaattctgtaatagatcagccttccaagtatggagaa	6319		

QY 601 ctggataatgtcacaaatgttcagattctctcttcgtatgtcagaagaatgcaaaatgttcgc 660
 |||||||
 DB 5900 ctgataatgtcacaaatgttcagattctctcttcgtatgtcagaagaatgcaaaatgttcgc 5959
 QY 661 gaatttcgtcttcttcagatcacagaagaatgttccagaatctttggacaactgcgc 720
 |||||||
 DB 5960 gaatttcgtcttcttcagatcacagaagaatgttccagaatctttggacaactgcgc 6019
 QY 721 tcaatgtggcagaactctcaatgttcttcaatcacatgtgaaataattctattgtgaa 780
 |||||||
 DB 6020 tcaatgtggcagaactctcaatgttcttcaatcacatgtgaaataattctattgtgaa 6079
 QY 781 atagataagtagtctggagctctaaatgtctgtatcagtaaaaggccaatcacagat 840
 |||||||
 DB 6080 atagataagtagtctggagctctaaatgtctgtatcagtaaaaggccaatcacagat 6139
 QY 841 ggaagacagattgaaagaagaacaattcatgcatctgttaattgtagtactaaaga 900
 |||||||
 DB 6140 ggaagacagattgaaagaagaacaattcatgcatctgttaattgtagtactaaaga 6199
 QY 901 aggcagctcaaaagacacacacagtttcaacaataaccatactatgatagacagactct 960
 |||||||
 DB 6200 aggcagctcaaaagacacacacagtttcaacaataaccatactatgatagacagactct 6259
 QY 961 gaaggaatcttactacgttgcataattgtcaatagatcacagctttcaagratgagaa 1020
 |||||||
 DB 6260 gaaggaatcttactacgttgcataattgtcaatagatcacagctttcaagratgagaa 6319
 QY 1021 gtctcaaaagtcatatagctcacagaacaagatatcttcttccacgtggt 1080
 |||||||
 DB 6320 gtactcaaaagtcatatagctcacagaacaagatatcttcttccacgtggt 6379
 QY 1081 gggagtgcctgcctgcattactacataataaaacccatagagcttctgtttttg 1140
 |||||||
 DB 6380 gggagtgcctgcctgcattactacataataaaacccatagagcttctgtttttg 6439
 QY 1141 cctcttcttcttgagactggcgctgcattcatgtgaaatggccacttgcactgatca 1200
 |||||||
 DB 6440 cctcttcttcttgagactggcgctgcattcatgtgaaatggccacttgcactgatca 6499
 QY 1201 gccagaaggaactgtgctgcatgataatggagttggttcgaagtgcagtgaatac 1260
 |||||||
 DB 6500 gccagaaggaactgtgctgcatgataatggagttggttcgaagtgcagtgaatac 6559
 QY 1261 agtttaatgcacgacattatagctcctgcataatg-igaattgtcaataacgtttaaaaaa 1319
 |||||||
 DB 6560 agtttaatgcacgacattatagctcctgcataatg-igaattgtcaataacgtttaaaaaa 6619
 QY 1320 cggatattccctgctgtctgatacaacattacagtggttacagaacacccattcatggt 1379
 |||||||
 DB 6620 cggatattccctgctgtctgatacaacattacagtggttacagaacacccattcatggt 6679
 QY 1380 gtaaaagacactttaaagaagt 1401
 |||||||
 DB 6680 gtaaaagacactttaaagaagt 6701
 RESULT 4
 AAH20182
 ID AAH20182 standard; DNA: 12793 BP.
 XX
 AC AAH20182;
 XX
 DT 09-AUG-2001 (first entry)
 DE Human mutated spastin nucleotide sequence SEQ ID NO:15.
 XX
 XX Human; mouse: spastin; ARSACS: chromosome 13q11; identification:
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibres;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.

XX OS Homo sapiens.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 77..1156
 FT /tag= a
 FT /product= "mutated spastin"
 XX
 PN MO200129266-A2.
 XX
 PD 26-APR-2001.
 XX
 PE 20-OCT-2000; 2000WO-US29130.
 XX
 PR 20-OCT-1999; 99US-0160588.
 XX
 PA (UYMC-) UNIV MCGILL.
 PA (HOP1-) HOPITAL SAINT-Justine.
 XX
 PI Hudson TJ, Engert J, Richter A;
 XX
 XX WPI: 2001-308494/32.
 DR P-PSDB: AAB97823.
 DR
 XX
 PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 FT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -
 XX
 XX Claim 1; Page -: 76pp; English.
 PS
 CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes for
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with Spastin
 CC protein or spastin gene expression or activity. The present sequence
 CC encodes a mutated human spastin.
 CC N.B. The present sequence is not given in the present specification but
 CC is derived from the human spastin nucleotide sequence (AAH20174) as
 CC stated on page 14.
 XX
 SQ Sequence 12793 BP; 4163 A; 2257 C; 2487 G; 3886 T; 0 other;
 Query Match 99.2%; Score 1390; DB 22; Length 12793;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1401; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 1 gtagcagtaaaactagagcagctcccaagcgacacaaagccttagaagatatgcatcc 60
 |||||||
 DB 5300 gtagcagtaaaactagagcagctcccaagcgacacaaagccttagaagatatgcatcc 5359
 QY 61 aatgctgttttacaacactgtgcacagaattggcgagaagaataattgcacgcaga 120
 |||||||
 DB 5360 aatgctgttttacaacactgtgcacagaattggcgagaagaataattgcacgcaga 5419

QY 121 attaagagcctccttaataatgataatcctcttgaaagaagaatttgaaagagctcttcaa 180
 |||||
 Db 5420 attaagaagcctccttaataatgataatcctcttgaaagaagaatttgaaagagctcttcaa 5479
 QY 181 aatgcgtatgatgcaaaaggcgacagaatctgtttgtgttgatccctaaacagatcca 240
 |||||
 Db 5480 aatgcgtatgatgcaaaaggcgacagaatctgtttgtgttgatccctaaacagatcca 5539
 QY 241 gttgatagaatatttgaataagtggcccatctgcaaggccagcactttgtgtac 300
 |||||
 Db 5540 gttgatagaatatttgaataagtggcccatctgcaaggccagcactttgtgtac 5599
 QY 301 aacaacacgcatcttaacagaagaatgattgaggaattcagaatcttgaaaggcagc 360
 |||||
 Db 5600 aacaacacgcatcttaacagaagaatgattgaggaattcagaatcttgaaaggcagc 5659
 QY 361 aaagaggaaatccttaataaacttgacagatggaatagatccaattctgtatcat 420
 |||||
 Db 5660 aaagaggaaatccttaataaacttgacagatggaatagatccaattctgtatcat 5719
 QY 421 atcacagactgcccactcttattcttgcaatgacatccctgtgtattttgatccat 480
 |||||
 Db 5720 atcacagactgcccactcttattcttgcaatgacatccctgtgtattttgatccat 5779
 QY 481 gccagatatgacacaggggccacatccatagtcgccgagcagtgtttagaatttgat 540
 |||||
 Db 5780 gccagatatgacacaggggccacatccatagtcgccgagcagtgtttagaatttgat 5839
 QY 541 gcagattttagagacagatctcagatgttcagatcttattcttggaacccatttaa 600
 |||||
 Db 5840 gcagattttagagacagatctcagatgttcagatcttattcttggaacccatttaa 5899
 QY 601 ctggaataattgacaaatgttccaatttccctcttcgtaatgcaaaaatggcaaaagtctg 660
 |||||
 Db 5900 ctggaataattgacaaatgttccaatttccctcttcgtaatgcaaaaatggcaaaagtctg 5959
 QY 661 gaatttcgtctgttccagcatcagacagaatggtccagatcttttgacaacatgctgc 720
 |||||
 Db 5960 gaatttcgtctgttccagcatcagacagaatggtccagatcttttgacaacatgctgc 6019
 QY 721 tcagatggggcagaactctctaatttcttaatacacaatggaaaaatttatttggaa 780
 |||||
 Db 6020 tcagatggggcagaactctctaatttcttaatacacaatggaaaaatttatttggaa 6079
 QY 781 atagataagatgactgagctcttaaatgtgtatcttaagtaaaaggcaaaatccacagat 840
 |||||
 Db 6080 atagataagatgactgagctcttaaatgtgtatcttaagtaaaaggcaaaatccacagat 6139
 QY 841 gggagacagatgaaaaggaaacaattcagatcctgttaattgatatgttactaaag 900
 |||||
 Db 6140 gggagacagatgaaaaggaaacaattcagatcctgtgttaattgatatgttactaaag 6199
 QY 901 aggcagctcaaaagacataccaggttcaacaataactatactatgtaacttggagactc 960
 |||||
 Db 6200 aggcagctcaaaagacataccaggttcaacaataactatactatgtaacttggagactc 6259
 QY 961 gaaggagaatcttactacgtggtctaatttgytaatagatcagcgttttcaagtatgggaaa 1020
 |||||
 Db 6260 gaaggagaatcttactacgtggtctaatttgytaatagatcagcgttttcaagtatgggaaa 6319
 QY 1021 gtaacaaagtgatcatatcagctcaacaagaagaatattactctttccacagtggt 1080
 |||||
 Db 6320 gtaacaaagtgatcatatcagctcaacaagaagaatattactctttccacagtggt 6379
 QY 1081 ggaagtgcgtcgtcatctactcaaacataaaaaaaccacataggcctctgtttttg 1140
 |||||
 Db 6380 ggaagtgcgtcgtcatctactcaaacataaaaaaaccacataggcctctgtttttg 6439
 QY 1141 cctcttctttggaagctggcgtgcacatttcatgtgaatggcgaacttgcacgtgatca 1200
 |||||
 Db 6440 cctcttctttggaagctggcgtgcacatttcatgtgaatggcgaacttgcacgtgatca 6499

QY 1201 gccagaagaaacctgtgcgtgatgataatgaggtgtgttcgcgaatgactggaataac 1260
 |||||
 Db 6500 gccagaagaaacctgtgcgtgatgataatgaggtgtgtgttcgcgaatgactggaataac 6559
 QY 1261 agtttaatgacagcatataatagctcgtcatatg-tyaattgctaatacagttaaaaaa 1319
 |||||
 Db 6560 agtttaatgacagcatataatagctcgtcatatg-tyaattgctaatacagttaaaaaa 6619
 QY 1320 cggtaattccctggtcttgcgtccaacattacagtggttaagaagaccctattcatgtt 1379
 |||||
 Db 6620 cggtaattccctggtcttgcgtccaacattacagtggttaagaagaccctattcatgtt 6679
 QY 1380 gtaaaagcacacttaaaagaagt 1401
 |||||
 Db 6680 gtaaaagcacacttaaaagaagt 6701

RESULT 5

AAH20179 standard; DNA; 12793 BP.

AAH20179;

09-AUG-2001 (first entry)

Human mutated spastin nucleotide sequence SEQ ID NO:12.

KW Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.

OS Homo sapiens.

OS Synthetic.

PN MO200129266-A2.

PD 26-APR-2001.

PF 20-OCT-2000; 2000MO-US29130.

PR 20-OCT-1999; 99US-0160588.

PA (UYMC-) UNIV MCGILL.

PA (HOP1-) HOPITAL SAINT-JUSTINE.

PI Hudson TJ, Engert J, Richter A;

WPI; 2001-308494/32.

PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -

XX Claim 1; Page -: 76pp; English.

PS The present invention describes human and mouse spastin, and mutated

CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay

CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to

CC chromosome 13q11. (I) have neuroprotective activities and can be used in

CC gene therapy and as a spastin polypeptide agonists. (I), their fragments

CC or their complements can be useful for assaying the presence of a nucleic

CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the

CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or

CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (1). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with Spastin
 CC protein or spastin gene expression or activity. The present sequence
 CC encodes a mutated human spastin.
 CC N.B. The present sequence is not given in the present specification but
 CC is derived from the human spastin nucleotide sequence (AAH20174) as
 CC stated on page 14.

SQ Sequence 12793 BP; 4163 A; 2257 C; 2487 G; 3886 T; 0 other;

Query Match 99.1%; Score 1388.4; DB 22; Length 12793;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1400; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 gtagcaatgaactagagcagctcccaagcgacacaaagccttagaagatagcatcc 60
 Db 5300 gtagcaatgaactagagcagctcccaagcgacacaaagccttagaagatagcatcc 5359
 QY 61 aatgctgttttaacaacttggcacagaatttggcgcaagaagaatttaccagcgaga 120
 Db 5360 aatgctgttttaacaacttggcacagaatttggcgcaagaagaatttaccagcgaga 5419
 QY 121 attaagagcattcttaatgcatatccttctgaaaagaatgttgaagaacttcttcaa 180
 Db 5420 attaagagcattcttaatgcatatccttctgaaaagaatgttgaagaacttcttcaa 5479
 QY 181 aatgctgatgtgaaaaggcgacagaatctgtttgtgttgatccttagacagcatcca 240
 Db 5480 aatgctgatgtgaaaaggcgacagaatctgtttgtgttgatccttagacagcatcca 5539
 QY 241 gttagatgaatttgaatgaatgtggcccatctgcaaggcgacagccttctgtgtac 300
 Db 5540 gttagatgaatttgaatgaatgtggcccatctgcaaggcgacagccttctgtgtac 5599
 QY 301 aacaacagcattttagaagaatgtagttagaaggaaatcttgaagaagcgacag 360
 Db 5600 aacaacagcattttagaagaatgtagttagaaggaaatcttgaagaagcgacag 5659
 QY 361 aaagaagggaatccttataaactggacagatggaagaatcattctgtatcat 420
 Db 5660 aaagaagggaatccttataaactggacagatggaagaatcattctgtatcat 5719
 QY 421 atccagacgtcccatcttattcttggcaatgacatccgtgtattttgacctcat 480
 Db 5720 atccagacgtcccatcttattcttggcaatgacatccgtgtattttgacctcat 5779
 QY 481 gccagatatgcacaggggccacatccatagctccggacgcatgttagagatttgat 540
 Db 5780 gccagatatgcacaggggccacatccatagctccggacgcatgttagagatttgat 5839
 QY 541 gcagaattttaggacacagttctcagatgttctgacattctatcggggaaccatttaa 600
 Db 5840 gcagaattttaggacacagttctcagatgttctgacattctatcggggaaccatttaa 5899
 QY 601 ctggaataatgacaaatgttcaagattcccttctgtaattgcaagaatgtgcaaaatttcg 660
 Db 5900 ctggaataatgacaaatgttcaagattcccttctgtaattgcaagaatgtgcaaaatttcg 5959
 QY 661 gaatttcgtctgttcagacatcagacagaatgttccagaaatctttagacaactgcgc 720
 Db 5960 gaatttcgtctgttcagacatcagacagaatgttccagaaatctttagacaactgcgc 6019
 QY 721 tcagaatggggcagaactcttaatttcttaatacaatgcaaaaaattctatttgttaa 780
 Db 6020 tcagaatggggcagaactcttaatttcttaatacaatgcaaaaaattctatttgttaa 6079
 QY 781 atagataagagtaactgagccttaaatgtgtgtattctagtaagaagggaataatccagat 840

Db 6080 atagataagagtaactgagccttaaatgtgtgtattctagtaagaagggaataatccagat 6139
 QY 841 ggaacagattgaaaaagaaacaattcatgcatctgttaattgatagttacttaaaaag 900
 Db 6140 ggaacagattgaaaaagaaacaattcatgcatctgttaattgatagttacttaaaaag 6199
 QY 901 aggcgcctcaagaatccagttcaacaataaccttactatgtatgtatgtatgagatct 960
 Db 6200 aggcgcctcaagaatccagttcaacaataaccttactatgtatgtatgtatgagatct 6259
 QY 961 gaagaaatcttactacgtgtgtaatttgaatagatcagcgttcttcaagtagaaga 1020
 Db 6260 gaagaaatcttactacgtgtgtaatttgaatagatcagcgttcttcaagtagaaga 6319
 QY 1021 gtaataaagtgtcatatcagctcaagaacaaagabataactcttccacgtgtgt 1080
 Db 6320 gtaataaagtgtcatatcagctcaagaacaaagabataactcttccacgtgtgt 6379
 QY 1081 ggaatgactgtcctgtatctactacacactataaaacccataggccttctgttttg 1140
 Db 6380 ggaatgactgtcctgtatctactacacactataaaacccataggccttctgttttg 6439
 QY 1141 cctcttcttggagactggcgtgcatcttcaatgtgaatggcacttggacatgttca 1200
 Db 6440 cctcttcttggagactggcgtgcatcttcaatgtgaatggcacttggacatgttca 6499
 QY 1201 gccagaagaacactgtgacgtatgataatgagattgtgttcgaagtgcagtgaaatac 1260
 Db 6500 gccagaagaacactgtgacgtatgataatgagattgtgttcgaagtgcagtgaaatac 6559
 QY 1261 agttaatgacagcatatagctctgtcatgt- tgaattgtataatgaatlaaaaaa 1319
 Db 6560 agttaatgacagcatatagctctgtcatgt- tgaattgtataatgaatlaaaaaa 6619
 QY 1320 cggatctccctgtgtcgtatccaaatattcagtggttaacaagaacccctattcatgt 1379
 Db 6620 cggatctccctgtgtcgtatccaaatattcagtggttaacaagaacccctattcatgt 6679
 QY 1380 gtaaaagacatttaagaagt 1401
 Db 6680 gtaaaagacatttaagaagt 6701

RESULT 6
 AAH20175
 ID AAH20175 standard; DNA; 11493 BP.
 XX
 AC AAH20175;
 XX
 DT 09-AUG-2001 (first entry)
 XX
 DE Mouse spastin nucleotide sequence SEQ ID NO:3.
 XX
 KW Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
 XX
 OS Mus musculus.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..11493
 FT /*tag= a
 FT /product= "spastin"
 XX
 FN W0200129266-A2.
 XX
 PD 26-APR-2001.
 XX
 PF 20-OCT-2000; 2000WO-US29130.

XX 20-OCT-1999; 99US-0160588.
PR
XX
PA (UYMC-) UNIV MCGILL.
XX (HOPI-) HOPITAL SAINTE-JUSTINE.
XX
PI
XX Hudson TJ, Engert J, Richter A;
DR
DR WPI: 2001-308494/32.
DR P-PSDB; AAB97620.
XX
XX
XX New isolated polynucleotide, encoding spastin gene, and polypeptides,
PR useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -
XX
XX Claim 1; Fig 8; 76pp; English.
PS

CC The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Seguenay
CC (ARSAACS) gene sequences (1). The spastin gene has been mapped to
CC chromosome 13q11. (1) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (1), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (1) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (1) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (1). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with Spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes mouse spastin as given in the present invention.
XX
SQ Sequence 11493 BP; 3599 A; 2281 C; 2387 G; 3226 T; 0 other;

Query Match	83.8%	Score 1174	DB 22	Length 11493
Best Local Similarity	90.3%	Pred. No. 0		
Matches 126; Conservative	0	Mismatches 135	Indels 1	Gaps 1

OY	1	gtacgcagtaaaacttagagcagctcccaaacgcacacaaaccttagaagaatgatgcattc	60
Db	5224	gtacgcgtgtaaaacttctgtgcaataaccgaagaacataaagcattagaagaattagattaccc	5283
OY	61	aatgctctgttttaacaacacttggcacagaattctgggcagaagaanaaaattgaccagcaga	120
Db	5284	aacatctgtttcccaagctcctcaggtacagaaatttgggcagaagaanaaaactacacacaga	5343
OY	121	attaagagatcccttaagtacataccctctcggaaaagaanaattgtgaagaagcttcttcaa	180
Db	5344	attaagagcatcttcaatgcattcctctccgaaaaagaanaatgctcgaagaagcttcttcaa	5403
OY	181	aatgctcatgatgycaaagcgacagaaatctgttttggtttgaatctctagaacagatcca	240
Db	5404	aatgctcatgatgycaaagcgacagagaatctgtcttcttggtaacctagaacagatcct	5463
OY	241	gttgcataagatatattgatgaataagtggccccattgcaagggccagacacttggctgac	300
Db	5464	gtttgaccggaattatttgatgtaagtggccccactgcaagggccagacacttggcttacc	5523
OY	301	aacaaacagccatttcaagaagaatgattgtatagaagaattcagaactcttgaaaaagcagc	360
Db	5524	aacaaacagccatttcaagaagaatgattgtatagaagaattcagaactcttgaaaaagcacc	5583

QY	361	aagagggaaatccttctaataactggaacagtagatggatagggttcaattctgtgtatcat	420
Db	5584	aaagaaaggaaatccctctgcaaaacaaaggacattlaagaaatcgatccaattccgtttatcat	5633
QY	421	atccagagctcccatctctttatctctgtggaatgaaatccgtgtatctttgtatcccat	480
Db	5644	attacagacttgcctctctttattcttctgtggaatgacatccctggtattctttgtatcccat	5703
QY	481	gccaagatatgcacgaagggccacaatccatagttcccggaacgaatgttaagagatttgat	540
Db	5704	gccaagatatgcacgaagggccacaatccatagttgagcctctggaacgaatgttaagagatttgat	5763
QY	541	gcaagatttgaagacagcttccatagatgtgtctgtgaattcttaactcttggaacccatttaa	600
Db	5764	gcaagactttgaagaccgcgttccatagatgtgtctgtgaattcttactgttgggaacccatttaa	5823
QY	601	cttgataatttgcacaaatglttcagatttcctctctgttaatgbcgaatvtggcaaaagtctgc	660
Db	5824	cttgagaaattgttaacaaatgltttagatttcctctctgttaatgbcagatvtggcaaaagtctgc	5883
QY	661	gaaatttcgtctgtctcagatataagacagaatvtgtcccaaatcttttggacaacactgtgc	720
Db	5884	gaaatttcgtctgtctcagatataagacagaatvtgtcccaaatcttttggacaagtctgcg	5943
QY	721	tcaagatgggcagaaactctcaatgatttctctaactcaatggaanaaaattctatctgttga	780
Db	5944	tctgattgggcgaagactctcaatgatttctctcaaccacaaatggagagaaatatctattgtgaa	6003
QY	781	atagataagagtagctgagcctctaataatgtgtctgtattccaagtaagggccaaaatcacagat	840
Db	6004	atagataagagtagctgagcctctaataatgtgtctgtattccaagtaagggccaaaatcacagat	6063
QY	841	ggagacagattggaagaagaaacaattccatgcatactcgttaattgtagtgcgttactaaag	900
Db	6064	ggagacagattggaagaagaaaggaagcattcccaagcctctgtaattgtacagvtgtactaaag	6123
QY	901	aggcagctcaaaagcatalaccagttcaacaataactatactatactatgatacttgaagactct	960
Db	6124	aggcagctcaaaagcatalaccagttcaacaataactatactatactatgatacttgaagactct	6183
QY	961	gaaggaaatcttactatctgtgcgttaatttgaatagaatgaggtcttcaagtaattggagaa	1020
Db	6184	gaaggaaatcttgcaccacaatgtgcatactgtglaaagatccagagatttccagtaattggagaa	6243
QY	1021	gtatctaanaagtgtgcataatcagcttccaaagaaacaaagatatctctttcccaagvtgt	1080
Db	6244	gtatccaagagttgtaataatacagcttccaaagaaacaaagatatcaccctttcccaagvtgt	6303
QY	1081	ggagtagctgcctgcatactactacaactataaaaaaaccacaaaggcctctctgtttttg	1140
Db	6304	ggagtagtagcctgcatactactacaactataaaaaaaccacaaaggcctctctgtctttg	6363
QY	1141	cctcttcttttggagacttggcgtgcacatttcaatgttlaatgycacacttgcacattgattatca	1200
Db	6364	cctctctcttggagaaaggcgtgcacatttcaatgttlaatgycacacttgcctctagattca	6423
QY	1201	gccaagaaagacactgtggtgtatgataatgagagttgtgtctcgaaagtgcacvtgataac	1260
Db	6424	gccaagaaagaaacttgtggtgtatgataatgagagttgtgtctcgaaagtgcacvtgataat	6483
QY	1261	agtttaatgaagaagataatagctcctgcgtatg-rgaatgtcctaatacagttaaaaaa	1319
Db	6484	agtttaatgaagaagataatagacacctgcgtatggtatcctaatacagttaaaaaa	6543
QY	1320	cggatattccctgttcttgcatacaatatacagvtgttacaanaaacccctataatgt	1379
Db	6544	cggatattccctgttcttgcatacaatatacagvtgttacaanaaacccatcatgtc	6603
QY	1380	gtaaagagacactttaagaagt 1401	
Db	6604	gtaaagagacacatttaagaagt 6625	

Db	6304	ggagtagcagccctgcattactcaatataataaaagccccaagagccttcctgtcttctg	6363
Qy	1141	ccctcttcttggagactggcctgcacattcatgtgaatgycacacttgcacattgatca	1200
Db	6364	ctctcctcttggagacagcgtgcacattcatgtgaatgycacacttgccttagatca	6423
Qy	1201	gccagaggaagaccgtgtgcgtgatgataagagtggtgtgttcggaagtgagcggaaatcac	1260
Db	6424	gccagaaagaaactctgtgcgtgatgataaaggggttgcgttcggaagtgagcggaaatcac	6483
Qy	1261	agtttaatgacagcattatagctccctgcacatg-tgaaattgctatacagttaaanaa	1319
Db	6484	agtttaatgacagcattatagcactctgtcattatggtgattctatccattcaattaaaaa	6543
Qy	1320	cggtaattccctgttcttgatccaacattatcagttgltacagaacaaccctatcatggt	1379
Db	6544	cggtaattccctgttcttgagcccaacattatcagtttaccagaacaacaccctatcatgc	6603
Qy	1380	gtaaagagacactttaagaagt 1401	
Db	6604	gtaaagagacacatttaagaagt 6625	

ID ABA28269 standard; DNA: 531 BP.
XX
AC ABA28269;
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #6735 for gene expression analysis in human heart cell sample.
XX
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX
PS Claim 1; SEQ ID No 6735; 530pp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart. The
XX present sequence is one such probe. The probes may be used for
XX predicting, measuring and displaying gene expression in samples derived
XX from the human heart via microarrays. By measuring gene expression, the
XX probes are useful for predicting, diagnosing, grading, staging,
XX monitoring and prognosing diseases of the human heart and vascular system
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
XX congenital heart disease.
XX
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 531 BP; 156 A; 108 C; 107 G; 160 T; 0 other;

Query Match 2.8%; Score 39.6; DB 22; Length 531;
Best Local Similarity 53.2%; Pred. No. 0.51;
Matches 84; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 1241 ttcgaaagtgcgtgaataacagtttaatacagcattatagctcctgcatagtgaatt 1300
DB 507 TCCCAAGTAGCTGGAACTACAGGTGATGCCACAGCCAGCAATTTTTTTTAAATT 448
QY 1301 gctaaacagtttaaaaaacggtatttccctggtctgataccaacattacagtggtaca 1360
DB 447 TTTTCATAGACAGACAGATCTCACTATGTTTCCCAAGTCTATAAACAATTATGATATAAAA 368
QY 1361 gaacacccctatcatgtgtgtaagacacttaaga 1398
DB 387 GAAAAAAGTAATCATCTGAAGTTAAGTCTTAATGA 350

RESULT 10

AAK08051/C
ID AAK08051 standard; DNA: 531 BP.
XX
AC AAK08051;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 8042.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX
PS Example 4; SEQ ID NO: 8042; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is one of the probes of the
XX invention.
SQ Sequence 531 BP; 156 A; 108 C; 107 G; 160 T; 0 other;

Query Match 2.8%; Score 39.6; DB 22; Length 531;
Best Local Similarity 53.2%; Pred. No. 0.51;
Matches 84; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 1241 ttcgaaagtgcgtgaataacagtttaatacagcattatagctcctgcatagtgaatt 1300
DB 507 TCCCAAGTAGCTGGAACTACAGGTGATGCCACAGCCAGCAATTTTTTTTAAATT 448
QY 1301 gctaaacagtttaaaaaacggtatttccctggtctgataccaacattacagtggtaca 1360
DB 447 TTTTCATAGACAGACAGATCTCACTATGTTTCCCAAGTCTATAAACAATTATGATATAAAA 388
QY 1361 gaacacccctatcatgtgtgtaagacacttaaga 1398
DB 387 GAAAAAAGTAATCATCTGAAGTTAAGTCTTAATGA 350

RESULT 11
AAK33925/C
ID AAK33925 standard; DNA: 531 BP.
XX
AC AAK33925;

OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI: 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID NO 8336; 654bp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 531 BP; 156 A; 108 C; 107 G; 160 T; 0 other;

Query Match 2.8%; Score 39.6; DB 22; Length 531;
Best Local Similarity 53.2%; Pred. No. 0.51;
Matches 84; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 1241 ttcgaaagtgcagtaaacagtttaacagcattacagtcctcgcatactgtaatt 1300
Db 507 TCCCAAGTACCTGGAAGTACAGGTGCGATGCCACCATGCCAGTCAATTTTATTATT 448
QY 1301 gctaatacagtttaaaaaacaggtattccctggtctcgtatccacattatcagtgta 1360
Db 447 TTTCATAGAGACAGAGAGTCTGATGTTTCCAGTCTCTAATAACAATTATGATAAAAA 388
QY 1361 gaacaccctatcatcgttgtaagagacacttaaga 1398
Db 387 GAAAAAAGTAAATCATCTGAGTAAAGTCTTTAATGA 350

RESULT 14
AAK87559/c
ID AAK87559 standard; DNA; 10119 BP.
XX
AC AAK87559;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human Immune/haematopoietic antigen genomic sequence SEQ ID NO:42371.
XX
XX Human; Immune; haematopoietic; Immune/haematopoietic antigen; cancer;
KW Cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.

XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227809.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251858.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI: 2001-483426/52.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 useful for preventing, diagnosing and/or treating cancers and

PT metastasis -
 XX
 PS Disclosure: SEQ ID NO 42371; 3071bp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
 CC represent sequences used in the exemplification of the present invention.
 CC
 XX
 SQ Sequence 10119 BP; 2866 A; 2254 C; 2160 G; 2839 T; 0 other;
 Query Match 2.88; Score 39.6; DB 22; Length 10119;
 Best Local Similarity 53.28; Pred. No. 1.9;
 Matches 84; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
 Qy 1241 ttggaagtgaactgaataacagtttaacagacattatagcttcgcatatgtaatt 1300
 Db 1842 TCCCAAGTAGCTGACACACAGCTGTCATGCCATGCCCATCTTTTTTTAATT 1783
 Qy 1301 gctaatacagtttaaaaaacggtattccctggtttcgtatccaacattacagtttaca 1360
 Db 1782 TTTTCATAGACAGACAGCTCTACTATGTTTCCACAGCTTAATAACATTATGATGAATAAA 1723
 Qy 1361 gaacacccctattcatgttgcataaaggacatttaaga 1398
 Db 1722 GAAAAAAGTAATCATCTGGAAGTTAAGCTTTAATGA 1685
 RESULT 15
 ID AAK73076 standard; DNA; 12247 BP.
 XX
 AC AAK73076;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27888.
 XX
 KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
 KW cytostatic; gene therapy; vaccine; metastasis; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200157182-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01354.
 XX
 XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 06:32:38 ; Search time 108.82 Seconds
(without alignments)
3162.402 Million cell updates/sec

Title: US-09-693-205-7_COPY_5300_6700

Perfect score: 1401
Sequence: 1 gtagagcagtaaacctagagc.....aaagacacttaagaagt 1401

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/prodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/prodata/1/ina/6A.COMB.seq:*
3: /cgn2_6/prodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/prodata/1/ina/6A.COMB.seq:*
5: /cgn2_6/prodata/1/ina/6A.COMB.seq:*
6: /cgn2_6/prodata/1/ina/6A.COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43.2	3.1	7218	1	US-08-232-463-14
2	35	2.5	87350	3	US-08-781-891-79
3	34.2	2.4	423	1	US-08-470-179-142
4	33.8	2.4	305	3	US-08-480-640A-225
5	33.8	2.4	305	4	US-08-686-968C-225
6	33.8	2.4	305	4	US-08-488-237A-225
7	33.8	2.4	2651	1	US-08-462-949-1
8	33.8	2.4	2651	1	US-08-023-764B-1
9	33.6	2.4	4203	2	US-08-866-757-1
10	33.6	2.4	1007	3	US-08-153-593-1
11	32.8	2.3	1007	3	US-08-924-747-3
12	32.8	2.3	1007	4	US-09-247-373B-3
13	32.8	2.3	1007	4	US-09-296-715-3
14	32.6	2.3	1750	3	US-09-120-365-90
15	32.6	2.3	1750	4	US-09-515-039-90
16	32.2	2.3	8920	2	US-08-446-855A-1
17	32.2	2.3	8920	2	US-08-446-855A-1
18	32.2	2.3	8920	2	US-08-446-855A-1
19	32.2	2.3	8920	2	US-08-446-855A-1
20	32.2	2.3	8920	2	US-08-446-855A-1
21	32.2	2.3	8920	2	US-08-446-855A-1
22	32.2	2.3	8920	2	US-08-446-855A-1
23	32.2	2.3	8920	2	US-08-446-855A-1
24	32.2	2.3	8920	2	US-08-446-855A-1
25	32.2	2.3	8920	2	US-08-446-855A-1
26	32.2	2.3	8920	2	US-08-446-855A-1
27	32.2	2.3	8920	2	US-08-446-855A-1

28	31.8	2.3	3499	2	US-08-658-665-52	Sequence 52, Appl
29	31.8	2.3	3499	4	US-09-085-773-52	Sequence 52, Appl
30	31.8	2.3	7859	1	US-07-854-596B-4	Sequence 4, Appl
31	31.8	2.3	7859	2	US-08-450-905B-15	Sequence 15, Appl
32	31.8	2.3	7859	3	US-07-982-759F-15	Sequence 15, Appl
33	31.4	2.2	2549	1	US-08-470-720-2	Sequence 2, Appl
34	31.4	2.2	4964	1	US-08-470-720-5	Sequence 5, Appl
35	31.2	2.2	1361	1	US-08-118-469A-4	Sequence 4, Appl
36	31.2	2.2	1361	1	US-08-909-119-4	Sequence 4, Appl
37	31.2	2.2	1400	4	US-09-117-257-10	Sequence 10, Appl
38	31.2	2.2	1400	4	US-08-945-476-10	Sequence 10, Appl
39	31.2	2.2	1400	4	US-09-489-352-10	Sequence 10, Appl
40	31.2	2.2	1498	1	US-07-965-668A-1	Sequence 1, Appl
41	31.2	2.2	1498	2	US-08-950-433-1	Sequence 1, Appl
42	31.2	2.2	1498	3	US-09-186-287-1	Sequence 1, Appl
43	31.2	2.2	2653	2	US-08-589-711-1	Sequence 1, Appl
44	31.2	2.2	2653	4	US-09-221-938-1	Sequence 1, Appl
45	31.2	2.2	2653	4	US-08-945-476-7	Sequence 7, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEFFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpc-Fls
US-08-232-463-14

	Query Match	Similarity	3.1%;	Score	43.2;	DB	1,	Length	7218;	
	Best Local	Similarity	5.8%;	Pred.	No.	0.016;				
	Matches	21;	Conservative	188;	Mismatches	151;	Indels	0;	Gaps	0;
QY	43	ttagaagaatagcatgccacgtgctcgttttacacaactgycgcagaatttggcgaa	102							::: :::
Db	1479	'TTACTATCTAATGACAGTAAAGAGCTATAAGCAATTGGTACRRRRRRRRRRR	1420							
QY	103	gaaaaattgaccgcagcaaatlaagacatcccttaatgcalatcctctgaagaanaaty	162							
Db	1419	RR	1360							
QY	163	ttgaagagcttcctcaaaatgctgatgatagtgaaggcgacgaaatcgttgttgctt	222							
Db	1359	RR	1300							
QY	223	gacctagacagcatccagctgatagaaatatgtgataagtggccccatgycgaagy	282	::: :::	::: :::	::: :::	::: :::	::: :::	::: :::	
Db	1299	RR	1240							
QY	283	ccagcacttgtygtacacaacacgacatltacagaagaatgatltagagaaattcg	342							
Db	1239	RR	1180							
QY	343	aactctggaanaagcgacgaagaaggaataccttataaaccgycgacagtatgaaataga	402							
Db	1179	RR	1120							

RESULT 2
 US-08-781-891-79/c
 : Sequence 79, Application US/08781891
 : Patent No. 6090620
 : GENERAL INFORMATION:
 : APPLICANT: Fu, Ying-Hui
 : APPLICANT: Yu, Chang-En
 : APPLICANT: Oshima, Junko
 : APPLICANT: Mulligan, John T.
 : APPLICANT: Schellenberg, Gerald D.
 : TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
 : NUMBER OF INVENTION: WERNER'S SYNDROME
 : NUMBER OF SEQUENCES: 209
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: SEED and BERRY LLP
 : STREET: 6300 Columbia Center, 701 Fifth Avenue
 : CITY: Seattle
 : STATE: Washington
 : COUNTRY: USA
 : ZIP: 98104-7092
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/781,891
 : FILING DATE: 27-DEC-1996
 : CLASSIFICATION: 800
 : ATTORNEY/AGENT INFORMATION:
 : NAME: No. 6090620tenburg Ph.D., Carol
 : REGISTRATION NUMBER: 39,317
 : REFERENCE/DOCKET NUMBER: 240052.419
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (206) 622-4900
 : TELEFAX: (206) 662-6031
 : INFORMATION FOR SEQ ID NO: 79:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 87350 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : US-08-781-891-79

Query Match	Similarity	2.5%	Score 35	DB 3	Length 87350
Best Local	Similarity 51.6%		Pred. Num. 8.4		
Matches 80	Conservative 0	Mismatches 75	Indels 0	Gaps 0	
Oy	819	aglaaaggccaacatcacagatcgtgacacgatatggaagaacatltcalgcgtcgt	878		
Db	49881	ACTGAGGGGGAAAACCGAAGAAAGATCTCGAATATAAAGAAAAAATTTTAAGCTTC	49822		
Oy	879	aatgatcgtgttacttaaaaagaagcagctcaagaacatcacagltccaacaataccta	938		
Db	49821	AGTTAATAGTGTGCTCCCTTGTTGGTAAAAATAAATAATGTATATATAAATAAAAAAAGAAAT	49762		
Oy	939	tactatgatacttgagagactcgaagaacatcta	973		
Db	49761	TTTTTAATCTAATAAAACGTTTAAATATATTTTAA	49727		

```

1      RESULT      3
2      US-08-470-179-142
3      Sequence 142. Application US/08470179
4      Patent No. 5645594
5      GENERAL INFORMATION:
6      APPLICANT: Huang Ph.D, Wai Mun
7      TITLE OF INVENTION: Method and Compositions for
8      TITLE OF INVENTION: Identification of Species in a Sample
9      NUMBER OF SEQUENCES: 207
10     CORRESPONDENCE ADDRESS:
11     ADDRESSEE: Trask, Britt and Rossa
12     STREET: P.O. Box 2550
13     CITY: Salt Lake City
14     STATE: Utah
15     COUNTRY: USA
16     ZIP: 84110
17     COMPUTER READABLE FORM:
18     MEDIUM TYPE: Floppy disk
19     COMPUTER: IBM PC compatible
20     OPERATING SYSTEM: PC-DOS/MS-DOS
21     SOFTWARE: Patentln Release #1.0, Version #1.30
22     CURRENT APPLICATION DATA:
23     APPLICATION NUMBER: US/08/470.179
24     FILING DATE:
25     CLASSIFICATION: 435
26     ATTORNEY/AGENT INFORMATION:
27     NAME: Sweigert Ph.D, Susan E.
28     REGISTRATION NUMBER: 36,289
29     REFERENCE/DOCKET NUMBER: 2601
30     TELECOMMUNICATION INFORMATION:
31     TELEPHONE: 801-532-1922
32     TELEFAX: 801-531-9168
33     INFORMATION FOR SEQ ID NO: 142:
34     SEQUENCE CHARACTERISTICS:
35     LENGTH: 423 base pairs
36     TYPE: nucleic acid
37     STRANDEDNESS: double
38     TOPOLOGY: not relevant
39     MOLECULE TYPE: DNA (genomic)
40     HYPOTHETICAL: NO
41     ANTI-SENSE: NO
42     ORIGINAL SOURCE:
43     ORGANISM: Mycoplasma arthritidis
44     JS-08-470-179-142

```

[illegible]

LENGTH: 305 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
US-08-488-237A-225

Query Match 2.4%; Score 33.8; DB 4; Length 305;
Best Local Similarity 48.2%; Pred. No. 2;
Matches 95; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 844 gacagatgaaagaagaacatttcacatcgtatgtagtgaatgaagaag 903
DB 280 GAATCATAGAAACAGCATATGATGTTCTTATCGAATATACACATTTCATGAGATA 221
QY 904 cagctcaagacataccagcttcaacaataacctatactatgatactgagactgaa 963
DB 220 AATAACCAAAATCCAAACATTACATATCTGTGCTGATTTCTTATTTATCTCTAAGT 161
QY 964 ggaatcttactcagcgctcaatttgtaataagtcagccttccaagatgagaagta 1023
DB 160 ACATACGATGATGATGCTTATTAAAAACACATCAACTATCATGAAATGATATATTA 101
QY 1024 tctaaagtgcatac 1040
DB 100 TTTAAATCTCATATC 84

RESULT 7
US-08-462-949-1

; Sequence 1, Application US/08462949
; Patent No. 5606022

; GENERAL INFORMATION:

; APPLICANT: Rasmussen, Beth Ann

; TITLE OF INVENTION: Cloning and Identification of a Two

; TITLE OF INVENTION: Component Signal Transducing Regulatory System from

; NUMBER OF SEQUENCES: 39

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Dady & Dady P.C.

; STREET: 805 Third Avenue

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10022

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/462,949

; FILING DATE:

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: 08/023,764

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Robinson, Joseph R.

; REGISTRATION NUMBER: 33,448

; REFERENCE/DOCKET NUMBER: 0646/1B024-US1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-527-7700

; TELEFAX: 201-753-6237

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2651 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-462-949-1

Query Match 2.4%; Score 33.8; DB 1; Length 2651;
Best Local Similarity 53.4%; Pred. No. 4.6;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 746 ttcttaatcacatggaataatttcatttggtaataagatagatcggagctctaa 805
DB 261 TTCTTATGAAAGTCAACATCTGATATATGATATGAGGCTTCTTCTTCTGA 320
QY 806 atgtgctgtaattcaagaaagcacaatacacagatgagacagattgaaagaaacaat 865
DB 321 GTTGCGCTTTTTCACAGTACGCTACATCGAAGAAATGCTGAGATGCCGTAAGAACAT 380
QY 866 ttcatgcatctgt 878
DB 381 TTAATACATCCGT 393

RESULT 8

US-08-023-764B-1

; Sequence 1, Application US/08023764B

; Patent No. 5679540

; GENERAL INFORMATION:

; APPLICANT: Rasmussen, Beth Ann

; TITLE OF INVENTION: Cloning and Identification of a Two

; TITLE OF INVENTION: Component Signal Transducing Regulatory System from

; NUMBER OF SEQUENCES: 39

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: American Cyanamid Company

; STREET: One Cyanamid Plaza

; CITY: Wayne

; STATE: New Jersey

; COUNTRY: United States

; ZIP: 07470-8426

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/023,764B

; FILING DATE: 26-FEB-1993

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Barnhard, Elizabeth M.

; REGISTRATION NUMBER: 31,088

; REFERENCE/DOCKET NUMBER: 31,658-00

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (201)831-3246

; TELEFAX: (201)831-3305

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2651 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-023-764B-1

Query Match 2.4%; Score 33.8; DB 1; Length 2651;
Best Local Similarity 53.4%; Pred. No. 4.6;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 746 ttcttaatcacatggaataatttcatttggtaataagatagatcggagctctaa 805

Db 261 TTGCTATGAAAGTCAACATCTGATATAGCATATATATGCGCTTTCCTTCTGCA 320
Qy 806 atggtcgtatctcgtatgaagggcaaatcacagatggaagatgtgaagaacaat 865
Db 321 GTTGGCTCTATTATCAAGTGAAGTACATCGAAGAAATGTTGAGATGCGTAAAGAACAT 380
Qy 866 ttcatgcatctgt 878
Db 381 TTAATACATCCGT 393

RESULT 9

US-08-866-757-1/c
; Sequence 1, Application US/08866757
; Patent No. 5838716
; GENERAL INFORMATION:
; APPLICANT: ELSHOURBAGY, NABIL A
; APPLICANT: LI, XIAOTONG
; APPLICANT: BERGSM, DEK J
; TITLE OF INVENTION: NOVEL 7TM RECEPTOR (H2CAA71)
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA
; STREET: P.O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/866,757
; FILING DATE: 30-MAY-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: PRESTIA, PAUL F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GH-70055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SRO ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4203 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-866-757-1

Query Match 2.4%; Score 33.6; DB 2; Length 4203;
Best Local Similarity 51.3%; Pred. No. 6.3;
Matches 78; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Qy 524 tgttagagattcgatgcagatttagcacagttctcagatgtctgacattatc 583
Db 2867 TGTCTGATTAAGATTAAGTGAACATTAAGTAAATTAAGCAAGATATTTCTTCACTATA 2808
Qy 584 tgggaaccatttaactgataatgcacatgltcagattcctccttcgtaatgag 643
Db 2807 GAAATACATTTTAAATTAATTTTAAATAATTAACAGCTGTCTTATCTTGTC 2748
Qy 644 aaatgcacaaagttcggaaattcgtctgt 675

Db 2747 TTAATACAAATAGTATTAATTTGCTTCTT 2716

RESULT 10

US-09-153-593-1/c
; Sequence 1, Application US/09153593A
; Patent No. 6174994
; GENERAL INFORMATION:
; APPLICANT: ELSHOURBAGY, NABIL A
; APPLICANT: LI, XIAOTONG
; APPLICANT: BERGSM, DEK J
; TITLE OF INVENTION: NOVEL 7TM RECEPTOR (H2CAA71)
; FILE REFERENCE: GH-70055-1
; CURRENT APPLICATION NUMBER: US/09/153,593A
; CURRENT FILING DATE: 1998-09-15
; EARLIER APPLICATION NUMBER: 08/866,757
; EARLIER FILING DATE: 1997-05-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 4203
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
; US-09-153-593-1

Query Match 2.4%; Score 33.6; DB 4; Length 4203;
Best Local Similarity 51.3%; Pred. No. 6.3;
Matches 78; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Qy 524 tgttagagattcgatgcagatttagcacagttctcagatgtctgacattatc 583
Db 2867 TGTCTGATTAAGATTAAGTGAACATTAAGTAAATTAAGCAAGATATTTCTTCACTATA 2808
Qy 584 tgggaaccatttaactgataatgcacatgltcagattcctccttcgtaatgag 643
Db 2807 GAAATACATTTTAAATTAATTTTAAATAATTAACAGCTGTCTTATCTTGTC 2748
Qy 644 aaatgcacaaagttcggaaattcgtctgt 675
Db 2747 TTAATACAAATAGTATTAATTTGCTTCTT 2716

RESULT 11

US-08-924-747-3
; Sequence 3, Application US/08924747
; Patent No. 6063570
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN
; APPLICANT: O'KEEFE, DANIEL
; TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
; SOFTWARE: MICROSOFT WORD VERSION 7.0A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/924,747
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA AXAMETHY
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CL-1108

TELECOMMUNICATION INFORMATION
TELEPHONE: 302-892-8112
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1007 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE: SOYBEAN
TISSUE TYPE: SEED
IMMEDIATE SOURCE:
CLONE: SSM.PK0026.G11
US-08-924-747-3

Query Match	2.38;	Score 32.8;	DB 3;	Length 1007;
Best Local Similarity	58.08;	Pred. No. 6;		
Matches	58;	Conservative	0;	Mismatches 42;
			Indels	0;
			Gaps	0

Qy 930 aataacctatactatgatactgaagcctctaagaacattactacgtgctaattg 989
| | | | | | | | | | | | | | |
Db 380 ACTGAACACTCAATTGGGAGAAAGTTGGCCCTCATGAAAACCTCCTTGGGCCCAAGTAT 439

Oy 990 caatagatcagcctttcaagatgagagaagtatctaa 1029
| | | | | | | | | | | | | | | | | | | |
Db 440 aattacagagcctttaaagcacctggaaaaagctattgaaa 479

RESULT 12
US-09-247-373B-3

; Sequence 3, Application US/09247373B
; Patent No. 6168954

```

; GENERAL INFORMATION:
;
; APPLICANT: MCGONIGLE, BRIAN
; APPLICANT: O'KEEFE, DANIEL

```

; TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE ENZYMES
 ; FILE REFERENCE: CL-1108-A
 ; CURRENT APPLICATION NUMBER: US/09/247,373B
 ;

PRIOR FILING DATE: 1997-09-05

```

; SOFTWARE: Microsoft Office 97
; SEQ ID NO 3
; LENGTH: 1007

```

```

; TYPE: DNA
; ORGANISM: SOYBEAN
ITS-09-247-373R-3

```

Query Match 2.38: Score 32,8: DB 4: Length 1007:

Best Local Similarity	58.0%;	Pred. No. 6;							
Matches	58;	Conservative	0;	Mismatches	42;	Indels	0;	Gaps	0.

QY 930 ataactatactatgatactgtagactctgaagaaatcctaactagtgcattttg 989
| | | ||| | | | | | | | | | |
Db 380 actgaactacattggggagaagtctggccctgatgaaaaaacttccttggcccagaagtat 439

```

QY      990 taatagatcagcgtttcaaglatgagaaaglatctaa 102
          | ||| | ||| | | ||| | |||
Db      440 aatagaagagcgtttaagcactgaaagctattgaa 479

```

RESULT 13

US-09-296-715-3
; Sequence 3, Application US/09296715
; Patent No. 6171839

```

; GENERAL INFORMATION:
;
; APPLICANT: MCGONIGLE, BRIAN

```

1

1 APPLICANT: O'KEEFE, DANIEL
 2
 3 TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE
 4
 5 TITLE OF INVENTION: ENZYMES
 6
 7 NUMBER OF SEQUENCES: 32
 8
 9 CORRESPONDENCE ADDRESS:
 10

ADDRESS: E.I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: UNITED STATES OF AMERICA
ZIP: 19898

```

;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: DISKETTE, 3.50 INCH
;
; COMPUTER: IBM PC COMPATIBLE
;

```

```

1  OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 2
2  SOFTWARE: MICROSOFT WORD VERSION 7.0A
3  CURRENT APPLICATION DATA:
4  ADDITION NUMBER: HC/00/206 71E

```

FILING DATE:
CLASSIFICATION:
AUTHOR/AGENCY INFORMATION:

NAME: FLOYD, LINDA AXAMETHY
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: CI-1108

TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-8112
TELEFAX: 302-773-0164

```

; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1007 base pairs

```

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;

```

```

; MOLECULE TYPE:  CDNA
;
; HYPOTHETICAL:  NO
; ANTI-SENSE:  NO
;

```

```
; ORIGINAL SOURCE:
; TISSUE TYPE: SOYBEAN
; IMMEDIATE SOURCE:
; GROWN: YES DT0006 011
```

US-09-296-715-3

Query Match	2.3%	Score 32.8;	DB 4;	Length 1007;
Best Local Similarity	58.0%;	Pred. No. 6;		
Matches 58: Conservative	0:	Mismatches 42:	Indels 0:	Gaps 0:

QY 930 aataacctatactatgatactgaggaactctgaaggaatcttactacgtygctaattg 989

Db 380 ACTGAACATACATTGGGGAGAAAGTTGGCCCTGATGAAAAACCTTCCTTGGGCCCAAGTAT 439

Qy 990 taataagatcagcctttcaagtatggaataaagtatctaaa 1029

Db 440 AATTAGAGAGGCTTTAAAGCACTGGAAAAAGCTATTGAAA 479

RESULT 14
US-09-120-365-90

```

; sequence 30, Application 05/05120505
; Patent No. 6103514
; GENERAL INFORMATION:
; ADDITIONAL INFORMATION:

```

INVENTOR: MARGALIT, Shmuel
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-144749
; CURRENT APPLICATION NUMBER: US/09/130 365

; CURRENT FILING DATE: 1998-07-22
 ; EARLIER APPLICATION NUMBER: JP 9-333 474
 ; EARLIER FILING DATE: 1997-11-18

```

; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO. 90

```

```

;
; LENGTH: 1750
; TYPE: DNA

```

```
; ORGANISM: Periplaneta americana
; FEATURE:
; OTHER INFORMATION: "n" bases represent undetermined nucleotides
US-09-120-365-90
```

```
Query Match          2.3%; Score 32.6; DB 3; Length 1750;
Best Local Similarity 49.1%; Pred. No. 8.5;
Matches 86; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
```

```
QY 753 tcaatggaagaaattcttcttctgtaataagataagagactgagctctaaatgtct 812
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 456 taacagattcaagtgtattgtgtacgaaccataaagtgcgtgaattgcgaaaaatgyc 515

QY 813 gtttcagtaaaaggcacaatccacagatggagacagattgaaaaggaacaattcatgc 872
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 516 gtttggtagacaaaaggacgaaggtcacacaatacacaatgtgtatcgtgtacaga 575

QY 873 atctgaatgtagtgttactaaagaagcgctcaagaacataccagttcaa 927
      |||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 576 atctgcaaatgattcctcgtgtttaaaatcgcatccagttcgatagaaatgaa 630
```

RESULT 15

```
US-09-515-039-90
; Sequence 90, Application US/09515039
; Patent No. 6214599
; GENERAL INFORMATION:
; APPLICANT: Natori, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-144749
; CURRENT APPLICATION NUMBER: US/09/515,039
; EARLIER FILING DATE: 2000-03-06
; EARLIER APPLICATION NUMBER: JP 9-333 474
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 90
; LENGTH: 1750
; TYPE: DNA
; ORGANISM: Periplaneta americana
; FEATURE:
; OTHER INFORMATION: "n" bases represent undetermined nucleotides
US-09-515-039-90
```

```
Query Match          2.3%; Score 32.6; DB 4; Length 1750;
Best Local Similarity 49.1%; Pred. No. 8.5;
Matches 86; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
```

```
QY 753 tcaatggaagaaattcttcttctgtaataagataagagactgagctctaaatgtct 812
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 456 taacagattcaagtgtattgtgtacgaaccataaagtgcgtgaattgcgaaaaatgyc 515

QY 813 gtttcagtaaaaggcacaatccacagatggagacagattgaaaaggaacaattcatgc 872
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 516 gtttggtagacaaaaggacgaaggtcacacaatacacaatgtgtatcgtgtacaga 575

QY 873 atctgaatgtagtgttactaaagaagcgctcaagaacataccagttcaa 927
      |||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 576 atctgcaaatgattcctcgtgtttaaaatcgcatccagttcgatagaaatgaa 630
```

Search completed: May 22, 2002, 06:34:06
Job time: 7551 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 04:26:55 ; Search time 3530.57 Seconds
(Without alignments)
2963.620 Million cell updates/sec

Title: US-09-693-205-7_COPY_1_500
Perfect score: 500
Sequence: 1 atgattcagaagacatc.....gcagaattgttaatacaaa 500

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenBank: 1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
------------	-------------	-------	--------	-------	-------------

1	500	100.0	12793	6	AX119931	Sequence
2	500	100.0	12793	9	AF193556	Human sapi
3	500	100.0	92693	9	AF157766	Human sapi
4	499	99.8	99819	2	AC079761	Human sapi
5	344	68.8	174140	2	AC069017	Human sapi
6	331.2	66.2	11493	10	AF193557	Mus muscu
7	328.6	65.7	11492	6	AX119933	Sequence
8	42.2	8.4	12029	3	AE001381	Plasmodu
9	42.2	8.4	116127	2	AC096291	Rattus no
10	41.6	8.3	110960	3	CER39A1A	Caenorhab
11	41.6	8.3	160624	9	AC060835	Human sapi
12	41.6	8.3	168982	2	AC020669	Human sapi
13	41.6	8.3	206235	2	AC107486	Human sapi
14	41.4	8.3	49623	3	AC087232	Caenorhab
15	40.2	8.0	39419	3	CER801C7	Caenorhab
16	40.2	8.0	161404	9	AP002376	Human sapi
17	39.8	8.0	157637	2	OSJN00122	Oryza sat
18	39.6	7.9	40242	8	SPAC1F5	S. pombe ch
19	39.6	7.9	95796	2	AC094350	Rattus no
20	39.6	7.9	169541	9	AL136178	Human DNA
21	39.6	7.9	175485	2	AC073802	Mus muscu
22	39.6	7.9	186496	2	AC020573	Human sapi
23	39.4	7.9	98913	8	AF069442	Arabidops
24	39.4	7.9	176749	9	AC096748	Human sapi
25	39.4	7.9	199200	2	AC109830	Human sapi
26	39.4	7.9	199615	8	ATCHR1V7	Human sapi
27	39.2	7.8	106601	9	AL357352	Human DNA
28	39.2	7.8	160750	9	HSJ177110	Human DNA
29	39.2	7.8	165660	2	AC027194	Human sapi
30	39.2	7.8	181298	2	AC008333	Drosophi1
31	39.2	7.8	184003	2	AC011058	Human sapi
32	39.2	7.8	93443	8	ATF2206	Arabidops
33	39.2	7.8	100028	8	ATF3C22	Arabidops
34	39.2	7.8	146802	9	AC025186	Human sapi
35	39.2	7.8	178113	2	AC099516	Human sapi
36	39.2	7.8	190044	9	AC021842	Human sapi
37	38.8	7.8	825	3	AF206632	Plasmodu
38	38.8	7.8	914	3	PRRESA	X55124 P. falciparu
39	38.8	7.8	21137	8	AC093617	Human sapi
40	38.8	7.8	104386	8	ATT32A17	Arabidops
41	38.8	7.8	131949	2	AP003341	Oryza sat
42	38.8	7.8	149313	2	AC023548	Human sapi
43	38.8	7.8	158084	8	AP003492	Oryza sat
44	38.8	7.8	199362	8	ATCHR1V24	Arabidops
45	38.4	7.7	2069	9	BC007641	Human sapi

ALIGNMENTS

RESULT 1
AX119931
LOCUS AX119931 12793 bp DNA
DEFINITION Sequence 1 from Patent WO0129266.
ACCESSION AX119931
VERSION AX119931.1 GI:14036678
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 (bases 1 to 12793)
Hudson,T.J., Engert,J. and Richter,A.
Identification of arsacs mutations and methods of use therefor
Patent: WO 0129266-A 1 26-APR-2001;
JOURNAL MCGILL UNIVERSITY (CA) ; Hopital Sainte-Justine (CA)
FEATURES
SOURCE
1..12793
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT	ORIGIN
4163 a 2256 c 2487 g 3887 t	

Query Match 100.0%; Score 500; DB 6; Length 12793;
 Best Local Similarity 100.0%; Pred. No. 6.9e-101;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgattcagaagaacacatgactgactgagctgagcttcaatccagaacgattgcaag 60
 Db 1 atgattcagaagaacacatgactgactgagctgagcttcaatccagaacgattgcaag 60
 QY 61 tcttatcagaagaatgataatcattctgctgagcagaagaattgattgtaagtga 120
 Db 61 tcttatcagaagaatgataatcattctgctgagcagaagaattgattgtaagtga 120
 QY 121 tccatttgaagaagaacagaataccaccatctgatttcaaggcttaagaatggttggaanaa 180
 Db 121 tccatttgaagaagaacagaataccaccatctgatttcaaggcttaagaatggttggaanaa 180
 QY 181 tcttatatacatttctcagaagattgacttatttgaatgagatgacacttaaccacag 240
 Db 181 tcttatatacatttctcagaagattgacttatttgaatgagatgacacttaaccacag 240
 QY 241 aactatactagaagaagatgatacattgctggaacatcattagactcagaatgattccatcgtt 300
 Db 241 aactatactagaagaagatgatacattgctggaacatcattagactcagaatgattccatcgtt 300
 QY 301 agcatttgaagaagatgatacattgctggaacatcattagactcagaatgattccatcgtt 360
 Db 301 agcatttgaagaagatgatacattgctggaacatcattagactcagaatgattccatcgtt 360
 QY 361 aaaaacttgagaaggtctgctccttaaaaaataagatgatacattacacacatccgattataa 420
 Db 361 aaaaacttgagaaggtctgctccttaaaaaataagatgatacattacacacatccgattataa 420
 QY 421 aaaaattatcattccaccattaccagatgctgttcttgcataatgagaagaatgattcatt 480
 Db 421 aaaaattatcattccaccattaccagatgctgttcttgcataatgagaagaatgattcatt 480
 QY 481 gcagaaatgtgtaacaa 500
 Db 481 gcagaaatgtgtaacaa 500
 ORGANISM human.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 12793)
 AUTHORS Engert,J.C., Berube,P., Mercier,J., Dore,C., Lepage,P., Ge,B.,
 Bouchard,J.P., Mathieu,J., Melancon,S.B., Schalling,M.,
 Lander,E.S., Morgan,K., Hudson,T.J. and Richter,A.
 ARSACS, a sporadic ataxia common in northeastern Quebec, is caused
 by mutations in a new gene encoding an 11.5-kb ORF
 Nat. Genet. 24 (2), 120-125 (2000)
 JOURNAL 20120709
 MEDLINE 2 (bases 1 to 12793)
 REFERENCE Engert,J.C., Berube,P., Dore,C., Lepage,P., Ge,B., and
 AUTHORS Richter,A.
 TITLE Direct Submission
 JOURNAL Submitted (08-OCT-1999) Genome Centre, Montreal General Hospital,
 1650 Cedar Ave., Montreal, QC H3G 1A4, Canada
 FEATURES
 source 1..12793
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="13"
 /map="between D13S232 and D13S292"

mRNA
 gene
 cds

1..12793
 /gene="SACS"
 /product="sacsin"
 1..12793
 /gene="SACS"
 77..11566
 /gene="SACS"
 /note="molecular chaperone"
 /codon_start=1
 /product="sacsin"
 /protein_id="AAE31262.1"
 /db_xref="GI:6907042"
 /translation="MNTFMPGRELIVQWYFPDENRNHPSVSLMKMKNLYIHFSEDL
 TLDPDEMLPRTILLEGQTCVELRLRIPSLVILIDSEEQNLDEFLADIVQKGFPL
 KIDASTOHLPIKKYHSP.PSAVIOIMEMPILOKONCNOTISILPHDKYRGLST
 TDSEKERRIIOELALFKRIHNSDQISSTYTLKGGKVLHHTAKLPADRLSISVLD
 SDEATIRLANMLKIDOLKTSCLKLVLKIOENAFYSHEEVTDLMVNLNLSLAKNE
 NPVNLMLPLKFIQISOEQMVAGLELFDIDIVLKDLPCENEGTYFPPSPDIL
 HSLRQIGLKNESLKEKDVQVAKIEALQVACPDQDVLKAKAKTLVLVKNHTLL
 OSSEGKMTLKKIKWVPACKERPNNPNSLWKGDLCLCAPPMQGVGAILGSSLP
 LVSIVHNLKALGIPTKPSLSAIVKHFRTIVDMYSKPSDEDDYQFOHLLLEIYGF
 MDHLNKGDSFALKFPWYWGKKRCPPLQAVYIKPLHIDLDLPYLIHNYKTYAKRHO
 LFRVGSIEELTSDHISMTVQIKLYLSDSDLSQESQNLALMLNLIHMLYSHQIPPS
 PNTVPPIHHSKNPSKLMKPIHBCCDIVDDINDLEDSVPIILVHDDIPMKTAE
 MLKVPCLSTRILIPENMGFEOGOREPLIVIRKNIILEEYPSVDIFELQNDANA
 TECSPLIDMRNMDIRENLDPGMACHGALMSFNNSQSDSDENVITLGLSLKRG
 ENDVGKREGESVYHIDPIIMSEFEMMDPNHNSKHKKNSGRTIMNWK
 QOKRLRFPMQKRPFDIVRCQQLPLVEAFYSTNGTLFRLSFTODEAKKSEVSTCY
 NTADIVSLDEFSLCGHRLLIFQSVKSMYLLKLTLETPSLAODVILIKKSCSK
 ALMTPLVSLKELAKLMKTCSSNKKLPDSEPKSCIIQTVBEFHHFRIRILQSP
 LFRGPDDPALFEMAKSGQSKPDELISQKTECTVETCTMLCTCMDEGALKFSLSEG
 RLGLVPCGAVGVQSLIDQKMTVPKHIEVFCYLPRLKGTGLPVHINCGFVNSR
 KETIMDTCGRMNTTMRHYIVKAYLQVLSVLDLATSGELMDTYAAVMPDDVDL
 DESVTCQGFTEIDAHGKGLTLPVSDGSTMKNVRLPDDSLIKRRDVGSAKRTF
 LKLYLKTGSKNLCAVELPSSVKLFEEBAGCKQLLEWTESEKOFSESEVFPNQIEA
 KDPVYVXCHADIPREVAVKLGAVPKRKHALERYASVNCFTTLEQKRLSRTIS
 ILNAVSEKMLEKELLONADADKATECFEDFROHVDIDFDDKAPLOGPALCVYN
 NOPFTEDDVAGTIONIGKGTKEGNDPYTGQIGTFNSYHTHTDCPSPTSGNDIICDP
 HARYARGATISISGRMRDADAPRQFSPVLDLITGTHKRDCTMRPRLNAEKA
 KVEISVSPASDRMVONLDLKLKSDGDELMLFHNHEKLSICEIDSTGALNLYSVK
 GKLTDDRLKRPKHASVIDSVTKRQLKDIPOQIITYTDDTDBSBNLTWILCNMS
 GFSSMEKVSYSIAHKNODITLFPFGVNAACITLHNYKKPRAFCFLPLSEGLPFP
 VNGHFLADSAARLMDRDNGVGRSDMNSLMTALAPAVELLQLKRPFGSDPT
 LSVLQMTPIHVVKDLTKPLFPFPVNRDIDPLQCLVRLVNLCEHDMKRLIPVRA
 PNIDGSDLSHAYITITMNSTSKTRPFPNLLQDELQHKHNDYITTRKYAENVY
 RLKHLILETFNVLVNCDETANLYHCLIDIDIVSVTPADISFLMTSESPPTNCH
 GKLPRLQOQNLKLFHSLKLLVDYCFERDAENIEVEGLDLTLTSLDVTGPKARK
 FLTYNHELIPRDLFPMNTLYLKSYNLILNCAKAFEDISFEDLLSVLPRTKYS
 CTWKONFESMRLKNAHFISESVKDEQEPDLPFIDVVDVLTLDQMLPRTKFLV
 SANQIVPBGVDLPLISIMHIAVFPNAPQOSVHNALMKACCTOLAIKIKOSKQAPV
 LLSCHRTANISFPTISIKALHIMQTSITFAEKIYVENDFACLLYFNCLNHLMSODI
 KILKSLPCTKYSISGRVYSIGKFGCTVLTLSISAEEKATKQSSSSAFLEKJLHLEL
 YEVIGCVPPVNDLEEVYKHLPLKLENSYDKMLHLYLKNRSLSAEELSIKQDLPEK
 LESLILIHANSRLKQAKHEDYDRTVVFEMVLEKFLIPNDPFRKLEOLIKRPNHVF
 MTSWVEFLNIGLKYLLISOOLQOPAKIEISVANTENWMSKETQNTQNLVLIHIFOR
 MDLSNPLKEKLSLPIELCPERAPAFIRPHPOYOVNGLPIIKNGAVQVNRFPQOC
 DVILQILMTSCPILPEKATPLSTKEQESDGLGPQBOLEOVYLMNLVNDLPDLVYINC
 RVLCNITITLLEWVKTRAKVLRISYIELSEKKEFRQLGVAFVAVEDWKLKLEBE
 VVNLNEYSDPFKPYLYKLPLELTFHQLFHLHLETDIISIKQVVEVLSRFXNSEKO
 LDPENKRTYKRVVSGFLRSLONSVSRVRSLEVRDLATVLPQDRLKSSLSLVEDE
 APHYKSRIOGNIGVOMVLDSCYLGKDHCFHKLMLFPOKRLPRLSLLILEQDE
 EMPKVCOPACSLQIOGRIOLILSSQPIGLIIRIMKHENDNAFLANREKIRCKRLR
 EGLKVSCEPEKLOTTLTKYKGNPIPHSRSEFPAFLKRGNAVILLYIQHSSKRIINFL
 ALAMLTKSATDNLISDTSTYLIMAGNDYIRIEKIDSLGVAKIDSSEPSKLEPMGET
 PIPAEIHVTLMDPMNVFYPGEVGLVLAEGDDIGSYQPTVYVAVIIOVEVEREAD
 NSSFLKTIYQIDIGSEYKIVSLDLYKFSRPRESSOSRDSAPSTSPTEFLTPCLR

repeat_region	/note="Alusg repeat: matches 9. .301 of consensus" 38703. .39008
repeat_region	/note="Alusg repeat: matches 1. .306 of consensus" 39790. .40093
repeat_region	/note="Alusx repeat: matches 1. .304 of consensus" 40126. .40416
repeat_region	/note="Alusg repeat: matches 1. .292 of consensus" 40444. .40733
misc_feature	/note="Alusg repeat: matches 1. .292 of consensus" 41322. .41405
repeat_region	/note="Single clone region. Assembly confirmed by restriction digest data." 41541. .41788
repeat_region	/note="Alusg repeat: matches 1. .248 of consensus" 44790. .45101
repeat_region	/note="Alusg repeat: matches 1. .313 of consensus" 45261. .45312
repeat_region	/note="13 copies 4 mer tgtt 88% conserved" 45899. .46206
repeat_region	/note="Alut repeat: matches 1. .307 of consensus" 46754. .47052
repeat_region	/note="Aluv repeat: matches 1. .298 of consensus" 47067. .47365
repeat_region	/note="Aluv repeat: matches 1. .299 of consensus" 47477. .47873
repeat_region	/note="L1A10 repeat: matches 5950. .6322 of consensus" 47889. .48229
repeat_region	/note="Alusx repeat: matches 1. .312 of consensus" 49168. .49212
repeat_region	/note="Alu repeat: matches 85. .126 of consensus" 49620. .49693
repeat_region	/note="L2 repeat: matches 1685. .1757 of consensus" 50704. .51032
misc_feature	/note="Alusgc repeat: matches 1. .308 of consensus" 52204. .53009
repeat_region	/note="Cpg island" evidence=not_experimental 53978. .54137
repeat_region	/note="Alusx repeat: matches 1. .160 of consensus" 54179. .54511
repeat_region	/note="L1M6 repeat: matches 5822. .6172 of consensus" 5565. .55949
repeat_region	/note="Alusx repeat: matches 1. .305 of consensus" 57331. .57390
repeat_region	/note="30 copies 2 mer ga 75% conserved" 57357. .57392
repeat_region	/note="9 copies 4 mer gaga 91% conserved" 57753. .57930
repeat_region	/note="MIR repeat: matches 82. .262 of consensus" 58260. .58389
repeat_region	/note="MIR repeat: matches 2. .153 of consensus" 58564. .58611
repeat_region	/note="24 copies 2 mer ca 93% conserved" 59350. .59533
repeat_region	/note="Alusg repeat: matches 129. .313 of consensus" 59922. .60223
repeat_region	/note="Aluho repeat: matches 85. .299 of consensus" 61036. .61144
repeat_region	/note="L2 repeat: matches 2581. .2696 of consensus" 62008. .62187
repeat_region	/note="TIGER1 repeat: matches 2238. .2418 of consensus" 62188. .62316
repeat_region	/note="Aluv repeat: matches 1. .129 of consensus" 62330. .62363
repeat_region	/note="Alu repeat: matches 261. .294 of consensus" 62362. .62565
repeat_region	/note="TIGER1 repeat: matches 1586. .1787 of consensus" 62566. .62865
repeat_region	/note="Alusg repeat: matches 1. .302 of consensus" 62866. .64385
repeat_region	/note="TIGER1 repeat: matches 46. .1586 of consensus" 64386. .64694
repeat_region	/note="Aluv repeat: matches 1. .306 of consensus"

	repeat_L_region	64695..64713	/note="TIGER1 repeat: matches 29. .46 of consensus"
	repeat_L_region	65068..65395	/note="L1Pb2 repeat: matches 5405. .5733 of consensus"
	repeat_L_region	65396..65569	/note="AluY repeat: matches 136. .309 of consensus"
	repeat_L_region	65571..65640	/note="L1Pb2 repeat: matches 5728. .5791 of consensus"
	repeat_L_region	65696..65717	/note="11 copies 2 mer ta 100% conserved"
	repeat_L_region	65725..66096	/note="L1Pb2 repeat: matches 5789. .6155 of consensus"
	repeat_L_region	66371..66410	/note="10 copies 4 mer ttgt 82% conserved"
	repeat_L_region	67566..67886	/note="AluJb repeat: matches 1. .299 of consensus"
	repeat_L_region	69748..69930	/note="MIR repeat: matches 6. .248 of consensus"
	repeat_L_region	70957..71267	/note="AluI repeat: matches 1. .311 of consensus"
	repeat_L_region	71279..71413	/note="MER1B repeat: matches 548. .680 of consensus"
	repeat_L_region	71411..71737	/note="MER1A repeat: matches 47. .485 of consensus"
	repeat_L_region	71780..72075	/note="AlusX repeat: matches 1. .295 of consensus"
	repeat_L_region	72145..72256	/note="MER11-internal repeat: matches 42. .175 of consensus"
	repeat_L_region	72454..72865	/note="MER11-internal repeat: matches 332. .739 of consensus"
	repeat_L_region	72873..73249	/note="MER11-internal repeat: matches 883. .1261 of consensus"
	Query Match	100.0%; Score 500; DB 9; Length 92693;	
	Best Local Similarity	100.0%; Pred. No. 5.1e-101;	
	Matches 500; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
OY	1 atgatttacagaagaccatgctacgcagtgcagttcctaataccagaacgaattgcaag	60	
Db	18909 ATGATTTCAGAGAACCATGTACCTCAGCTGCAGCTTTAAATCCAGAACAATTGGCACG	18850	
OY	61 tcttatccaaggaaagtaatgatalacatlcttgcgcctggcgagagaatgatgttcaa199la	120	
Db	18849 TCTTATCAAGGAAGTAATGAATACTCTCGGCTGCAGAGAAATTGATTCATCGGTA	18790	
OY	121 tccatttgataaaaacagaataatcaccatctgtttcatalgagcttaaga199tttggaaaaa	180	
Db	18789 TCCATTATGATGAAAAACAATAACACCCTACTGTTCATGCTTAAGATGGTTTGAAAAA	18730	
OY	181 tcttatatacatttttagagagatttgaacttatattgatgagatgacactatcccag	240	
Db	18729 TCTTATATFACATTTTTTAGAGAGATTGTGACTTATTGTATGATGATGCCATTATCCCAG	18670	
OY	241 aaclactagaagaagtlcagacalgtlctggaactla199actcagatbtcalc9lt	300	
Db	18669 AACCTACTAGAGGAAGTCAAGACATGTGTGAACCTATTAGACTCAGAGATTCATCGTT	18610	
OY	301 agtcaatttagacgatgaatctgaaagcagcttcagaat199ttaagacagacattgtaca	360	
Db	18609 AGTCATTTTAGCGATGTGATCTGAAGCAAGCTTCCAAATTTTGAAGACAGATGTGACA	18550	
OY	361 aaaacttggagaggttgttccttaaaaaaat199atgcatcattatACAACTCCGCTTATTAA	420	
Db	18549 AAAAATTTGGAGGGTTGTCTTAAAAAATTTAGATGCACTTATACAACTCCGCTTATTAA	18490	
OY	421 aaaatatatcatcaccattaccaagtlcgtt199tttcagataa199gaagaatgccatt	480	
Db	18489 AAAAATATFTTCATTCACCATTAACCAAGTGTCTTTTTCAGAGATAATGGAAGATGCCATT	18430	
OY	481 gcaagaattgtgtaatcaaa	500	

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
DB 18429	GCAGAAATTGTATATCAAA	18410					
4	AC079761	Homo sapiens chromosome UNK clone RP11-143c17, *** SEQUENCING IN PROGRESS ***; 44 unordered pieces.	99819 bp	DNA	Linear	HTG 10-SEP-2000	
TITLE	AUTHORS	JOURNAL	REFERENCE	AUTHORS	JOURNAL		
The sequence of Homo sapiens clone	Waterston,R.H.	Unpublished	Submitted (10-SEP-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA	Waterston,R.H.	Direct Submission		
2 (bases 1 to 99819)	Waterston,R.H.	Unpublished	Submitted (10-SEP-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA	Waterston,R.H.	Direct Submission		
NOTE: This is a 'working draft' sequence. It currently consists of 44 contigs. The true order of the pieces is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.							
1	1137:	contig of 1137 bp in length					
1138	1237:	gap of unknown length					
1238	2538:	contig of 1301 bp in length					
2539	2638:	gap of unknown length					
2639	3976:	contig of 1338 bp in length					
3977	4076:	gap of unknown length					
4077	5355:	contig of 1279 bp in length					
5356	5455:	gap of unknown length					
5456	6757:	contig of 1302 bp in length					
6758	6857:	gap of unknown length					
6858	8570:	contig of 1713 bp in length					
8571	8670:	gap of unknown length					
8671	9925:	contig of 1255 bp in length					
9926	10025:	gap of unknown length					
10026	11426:	contig of 1401 bp in length					
11427	11526:	gap of unknown length					
11527	13266:	contig of 1740 bp in length					
13267	13366:	gap of unknown length					
13367	14794:	contig of 1428 bp in length					
14795	14894:	gap of unknown length					
14895	16054:	contig of 1160 bp in length					
16055	16154:	gap of unknown length					
16155	17395:	contig of 1241 bp in length					
17396	17495:	gap of unknown length					
17496	19287:	contig of 1792 bp in length					
19288	21294:	contig of 1907 bp in length					
21295	21394:	gap of unknown length					
21395	22944:	contig of 1550 bp in length					
22945	23044:	gap of unknown length					
23045	24421:	contig of 1377 bp in length					
24422	24521:	gap of unknown length					
24522	25870:	contig of 1349 bp in length					
25871	25970:	gap of unknown length					

*	25971	27230:	contig of 1260 bp in length
*	27231	27330:	gap of unknown length
*	27331	28778:	contig of 1448 bp in length
*	28779	28878:	gap of unknown length
*	28879	30993:	contig of 2015 bp in length
*	30894	30993:	gap of unknown length
*	30994	32460:	contig of 1467 bp in length
*	32461	32560:	gap of unknown length
*	33561	33984:	contig of 1424 bp in length
*	33985	34084:	gap of unknown length
*	34085	35285:	contig of 1201 bp in length
*	35286	35385:	gap of unknown length
*	35386	37184:	contig of 1799 bp in length
*	37185	37284:	gap of unknown length
*	37285	39172:	contig of 1888 bp in length
*	39173	39272:	gap of unknown length
*	39273	40874:	contig of 1602 bp in length
*	40875	40974:	gap of unknown length
*	40975	42893:	contig of 1919 bp in length
*	42894	42993:	gap of unknown length
*	42994	44384:	contig of 1391 bp in length
*	44385	44484:	gap of unknown length
*	44485	45989:	contig of 1515 bp in length
*	46000	46099:	gap of unknown length
*	46100	48689:	contig of 2570 bp in length
*	48670	48769:	gap of unknown length
*	48770	50798:	contig of 2029 bp in length
*	50799	50888:	gap of unknown length
*	50899	52809:	contig of 1911 bp in length
*	52810	52909:	gap of unknown length
*	52910	55127:	contig of 2218 bp in length
*	55128	55227:	gap of unknown length
*	55228	58087:	contig of 2860 bp in length
*	58088	58187:	gap of unknown length
*	58188	61004:	contig of 2817 bp in length
*	61005	61104:	gap of unknown length
*	61105	64185:	contig of 3081 bp in length
*	64186	64285:	gap of unknown length
*	64286	67105:	contig of 2820 bp in length
*	67106	67205:	gap of unknown length
*	67206	70837:	contig of 3632 bp in length
*	70838	70937:	gap of unknown length
*	70938	75837:	contig of 4900 bp in length
*	75838	75937:	gap of unknown length
*	75938	80452:	contig of 4515 bp in length
*	80453	80552:	gap of unknown length
*	80553	84661:	contig of 4109 bp in length
*	84662	84761:	gap of unknown length
*	84762	90542:	contig of 5781 bp in length
*	90543	90642:	gap of unknown length
*	90643	94348:	contig of 3706 bp in length
*	94349	94448:	gap of unknown length
*	94449	99819:	contig of 5371 bp in length
Location/Qualifiers			
1..99819			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
/chromosome="UNK"			
/clone="RP11-143G17"			
1..1137			
misc_feature	/note="assembly_name:Contig20"		
	1238..2538		
misc_feature	/note="assembly_name:Contig23"		
	2639..3976		
misc_feature	/note="assembly_name:Contig30"		
	4077..5355		
misc_feature	/note="assembly_name:Contig35"		
	5456..6757		
misc_feature	/note="assembly_name:Contig39"		
	6858..8570		
misc_feature	/note="assembly_name:Contig40"		
	8671..9925		
misc_feature	/note="assembly_name:Contig41"		

```
location/Qualifiers
1. 96819
  /organism="Homo sapiens"
  /DB_xref="taxon:9606"
  /Chromosome="XNK"
  /clone="RP11-143G17"
1. 1137
  /note="assembly_name:Contig120"
1238. 2538
  /note="assembly_name:Contig123"
2639. 3976
  /note="assembly_name:Contig130"
4077. 5355
  /note="assembly_name:Contig135"
5456. 6757
  /note="assembly_name:Contig139"
6858. 8570
  /note="assembly_name:Contig140"
8671. 9925
  /note="assembly_name:Contig141"
10000. 10000
```

```

misc_feature      10026..11426
                    /note="assembly_name:Contig42"
misc_feature      11527..13266
                    /note="assembly_name:Contig44"
misc_feature      13367..14794
                    /note="assembly_name:Contig47"
misc_feature      14895..16054
                    /note="assembly_name:Contig48"
misc_feature      16155..17395
                    /note="assembly_name:Contig49"
misc_feature      17496..19287
                    /note="assembly_name:Contig50"
misc_feature      19388..21294
                    /note="assembly_name:Contig51"
misc_feature      21395..22944
                    /note="assembly_name:Contig52"
misc_feature      23045..24421
                    /note="assembly_name:Contig53"
misc_feature      24522..25870
                    /note="assembly_name:Contig54"
misc_feature      25971..27230
                    /note="assembly_name:Contig55"
misc_feature      27331..28778
                    /note="assembly_name:Contig57"
misc_feature      28879..30893
                    /note="assembly_name:Contig58"
misc_feature      30994..32460
                    /note="assembly_name:Contig59"
misc_feature      32561..33984
                    /note="assembly_name:Contig60"
misc_feature      34085..35285
                    /note="assembly_name:Contig61"
misc_feature      35386..37184
                    /note="assembly_name:Contig62"
misc_feature      37285..39172
                    /note="assembly_name:Contig63"
misc_feature      39273..40874
                    /note="assembly_name:Contig64"
misc_feature      40975..42893
                    /note="assembly_name:Contig65"
misc_feature      42994..44384
                    /note="assembly_name:Contig66"
misc_feature      44485..45999
                    /note="assembly_name:Contig67"
misc_feature      46100..48669
                    /note="assembly_name:Contig68"
misc_feature      48770..50798
                    /note="assembly_name:Contig69"
misc_feature      50899..52809
                    /note="assembly_name:Contig70"
misc_feature      52910..55127
                    /note="assembly_name:Contig71"
misc_feature      55228..58087
                    /note="assembly_name:Contig72"
misc_feature      58188..61004
                    /note="assembly_name:Contig73"
misc_feature      61105..64185
                    /note="assembly_name:Contig74"
misc_feature      64286..67105
                    /note="assembly_name:Contig74"

Query Match      99.8%; Score 499; DB 2; Length 99819;
Best Local Similarity 100.0%; Pred. No. 8.3e-101;
Matches 499; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgattacaggaagcctgtactcagctgacgtcttaataatccagaagattgcaag 60
    |||
Db 48099 atgattacaggaagcctgtactcagctgacgtcttaataatccagaagattgcaag 48158
    |||

QY 61 tcttatcaaggaagtaataatcattcgcctgcagagaattgattcattcaatgta 120
    |||
Db 48159 tcttatcaaggaagtaataatcattcgcctgcagagaattgattcattcaatgta 48218
    |||

QY 121 tccatttgatgaagaacgaatccatcgtlctcatggttaagatggttgaaaaa 180
    |||

```

```

Db 48219 tccatttgatgaagaacgaatccatcgtlctcatggttaagatggttgaaaaa 48278
    |||
QY 181 tcttatcaatcttcttcagagattgacttatttgatgagatgccatccacag 240
    |||
Db 48279 tcttatcaatcttcttcagagattgacttatttgatgagatgccatccacag 48338
    |||

QY 241 aactatacagaaggaagtcagacatgltggaactcatactagaatccatcgt 300
    |||
Db 48339 aactatacagaaggaagtcagacatgltggaactcatactagaatccatcgt 48398
    |||

QY 301 agtcatctagaagatgaatctggaagcacagctccagaattttagagaatgta 360
    |||
Db 48399 agtcatctagaagatgaatctggaagcacagctccagaattttagagaatgta 48458
    |||

QY 361 aaacttgaggaggtgtgtcttcaaaaatagatgatacatcacatccgattaa 420
    |||
Db 48459 aaacttgaggaggtgtgtcttcaaaaatagatgatacatcacatccgattaa 48518
    |||

QY 421 aaataatcatcacatcacatcacatgctgttctgcagataatggaagatgcatt 480
    |||
Db 48519 aaataatcatcacatcacatcacatgctgttctgcagataatggaagatgcatt 48578
    |||

QY 481 gcagaatgtgtaatcaa 499
    |||
Db 48579 gcagaatgtgtaatcaa 48597
    |||

```

```

RESULT 5
AC069017      174140 bp      DNA      linear      HTG 12-SEP-2001
LOCUS      Mus musculus clone MGS3-342116, WORKING DRAFT SEQUENCE, 5 unordered
DEFINITION      pieces.
ACCESSION      AC069017.20 GI:15559167
VERSION      HTGS_PHASI; HTGS_DRAFT; HTGS_FULFOP.
KEYWORDS      house mouse.
SOURCE      Mus musculus
ORGANISM      Mus musculus
REFERENCE      1 (bases 1 to 174140)
AUTHORS      Metzger,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
Dederich,D., Thomas,S., Okwuonu,G., Carllock,C., Garner,T.,
Addison,S., Pace,A., Williams,G., Bonin,D., Brooks,A., Brown,J.,
Buhay,C., Bunac,C., Burkett,C., Chacko,D., Chen,G., Chen,Z.,
Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,
Gorrell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hognes,M.,
Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
Kovar,C., Liu,J., Liu,W., Louisseged,H., Lozano,R.J., Martin,R.,
Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Ogih,M., Parish,B.,
Perez,L., Relter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstein,G.,
Worley,K. and Gibbs,R.
JOURNAL      Direct Submission
REFERENCE      2 (bases 1 to 174140)
AUTHORS      Worley,K.C.
TITLE      Direct Submission
JOURNAL      Submitted (17-MAY-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 12, 2001 this sequence version replaced gi:14787161.
COMMENT      ----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: MGAO
Center clone name: MGS3-342116

```


ACCESSION AE001381 AE001382
VERSION AE001381.1 GI:3845124
KEYWORDS malaria parasite P. falciparum.
SOURCE Plasmodium falciparum
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 12029)
AUTHORS Gardner,M.J., Wetzel,H., Carucci,D.J., Cummings,L.M., Araujo,L., Koonin,E.V., Shalton,S., Mason,T., Yu,K., Fujii,C., Pederson,J., Shen,K., Jing,J., Aston,C., Lai,Z., Schwartz,D.C., Peters,M., Salzberg,S., Zhou,L., Sutton,G.G., Clayton,R., White,O., Smith,H.O., Fraser,C.M., Hoffman,S.L. et al.
TITLE Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum
JOURNAL Science 282 (5391), 1126-1132 (1998)
MEDLINE 99021743
REMARK Erratum: ([published erratum appears in Science 1998 Dec 4;282(5395):1827])
REFERENCE 2 (bases 1 to 12029)
AUTHORS Gardner,M.J.
TITLE Direct Submission
JOURNAL Submitted (02-NOV-1998) The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20814, USA
FEATURES
source
1..12029
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="2"
/gene="PFB0210C"
/gene="PFB0210C"
complement(105..1619)
/note="identified by sequence similarity; putative"
/codon_start=1
/product="monosaccharide transporter"
/protein_id="A0C71831.1"
/db_xref="GI:3845125"
/translation="MTKSSKDCISENEKKGKSGFSPSTFVKVLSACIASIFGQVSVLITKNITVEFEWCKGKDRKLNSTQSSFLASVIGAVJGCGSGVLVQFGRSLSLIYNFELVSLITSTHPTLIFARLISGEGIGLVAVPMVISEMTHDKKGAVGMHOLFIFGIFVAVMGLANGEGKADSTPLTSFAMLRMLFSPVISLIGIALVVEFEKEETPYEFKEKRIEESKNILKKIYEDNDVDEPLNAIKAEVEONESAKKNSLSLSALKIPISRYRVIIGCLSLGLOOFGINVASNSNELYKPELDSHITLISVYMTAVNPLMTPAIYIVIERKIGKTLINGCVCVLAIVPTAIANEINRNSNVKLISIVATEPMIISFAVSGIPVLMITVLEHMPSPETIKDSASLASLVNMCALIVFSDIITKKSPSLTIVFSVMSILTEFFIFEFIKETKGGEIGSPYITMEEROKHMTKSVV"
complement(3799..5049)
/gene="PFB0215C"
complement(3799..5049)
/note="identified by sequence similarity; putative"
/codon_start=1
/product="protein with Egl-1like 3'-5' exonucl. domain"
/protein_id="A0C71832.1"
/db_xref="GI:3845126"
/translation="MTKCKLNINIMKRRVRRNSYGYNNININEKHVYDNDKRNIRKISDCKSCRECVDEIKNGNTNLKDFNMKMGIDIEGICKYGVSTIQCYEDIYTFDIYKCNVYLFYIVIKDILECDIIRKTHDCRDCSLYQYNIHLNILDQVAYMLLNNNNTVYQIISYDLDLKYLFINNNHRIYFKHKTITDNYILKRPIMKEIISYAIQVYILKPLICILDFEIKRKKEOEKNKYVNDQNNKIKOEKDKSNLISGKNISSFNNDQYHFKETIIDIILSHSKRYVNOFIPNSHKDEKELOKMGILGMYVSCNNTKMYILKLMKRGVLIINYONKYEIGDIVAAVIVNFRNDYINGLTXDEKILILDAGKTIPIREYLVQNIQFLNEEKI"
5364..6843
/gene="PFB0220W"
join(5564..5681,5897..6843)
/note="identified by sequence similarity; putative"
/codon_start=1
/product="ubiquitinone biosynthesis methyltrans."
/protein_id="A0C71833.1"
/db_xref="GI:3845127"
/translation="MKITKLIINYNFRKITVNWGSIISKASHFSTQHGQYDKSERICNF

GFQVSEEEKSRILVNLFSNVCKKVDIMDMMSLVLHREKQDFVKELDILKYSHYN
IDQVYQHKDHSNNEKIQKKNENPSTNGYNNNSYVSDIPNYKILDLAGTGDIA
ERILEKSKFYLKNNOSIPDHI1STQYLYPHIIVCVNNDMLNKGKKAATLTDONL
TWLVQNAENLSEVSNSIDVITLSFGIRNETNIPQALKEHRLKFGGRFLCLESKY
OCHIFNIPYKFLNANVPIIGKVVANDMAKYVASIESIQFTLPDELAOLFHOANFN
ITYTMTMGIVSHSAVKLV"
complement(7198..8724)
/gene="PFB0225C"
complement(7198..8724)
/note="predicted by Glimmer"
/codon_start=1
/product="hypothetical protein"
/protein_id="A0C71834.1"
/db_xref="GI:3845128"
/translation="MGLPRKNNKVPYCYGDYRHSISLERINNIILKYSNIETCKN
KKKCDLELYILASVDNFKKKYETECCKLDGYINDEKRIKINEVKGRRKIDCTPN
NKKIFPVNHLINDDDLFRKKRKKKKKIKITLKNCKNDRKLNKNEKDDHVES
YTKKMLNNNNNNNITNMVSEFGYNKRRKYTNKCIINEENKMKDEQHIDAKN
INTNLHDDKNDKIRKHSITQPLYSNNKEKIIISKVNLKDESTNITKCONVVDN
IDKOKHREKRSKSIENNISIIITERYKSHDHCNLEENKEEKYNNLQDVITN
CNNDKVLLEYHHEKELNVQIINDIDKNEKKNNKKNKKKOKNKKNEKERNKQ
EKEKNKKKEKKNKKKEKSKKKEKKNKKKKEKKNKKKNGDVLKHYENNLO
DVELYEEKIINVTKKDELSYKKNKSEKDIYHILSEYSNTLQYTSFLDYKKNME
"
complement(9054..10994)
/gene="PFB0230C"
complement(9054..10994)
/gene="PFB0230C"
/note="predicted by Glimmer"
/codon_start=1
/product="hypothetical protein"
/protein_id="A0C71835.1"
/db_xref="GI:3845129"
/translation="MSTILNFVKEQKMMNTLHITKNFMENTLKYTEEIKHDKDINNLMR
RIHEEIKELISSNGKRYPMERIKIYFLMKKPKKEGFPSPNDVLKOSFKNKIK
NLOSIRKKNIOIKLOKYOILIKGYVNVCKKYQOENLLKSPFSNOSQYNTL
KSPHSRRKRIIFYPYTKLRKRLRIRISHFEDRYVIHKGPLYTKKKKTIYNKY
IYIYIYIYIYIFEMYSFTLEYSNSIFKRYTHKKRYKEIIDIILDNKLMLH
FKYKRYKRRKKKKLHISRRKDRNDLYCKKKKELIYTHLPLRLRKRILKS
SNYNYLNKEGENIINKEEENIILHDEENIILHDEENIILHDEENIILHDEENIILH
EENIILHDEENIILHDEENIILHDEENIILHDEENIILHDEENIILHDEENIILH
EENIILHDEENIILHDEENIILHDEENIILHDEENIILHDEENIILHDEENIILH
KVEERKQYINDEDEKENVSDINVECKKMKDEYKKNENNSIENENIILDSKQNIIL
DINKERLISKEKKKKISRIKTKTLEDNNDIKENENFNIYDEKNIGKKEEYIIEE
KNRHNVTIQDKAKMDNIDEO"
BASE COUNT 4304 a 1428 c 922 g 5375 t
ORIGIN
Query Match 8.4%; Score 42.2; DB 3; Length 12029;
Best Local Similarity 49.3%; Pred. No. 9.8;
Matches 110; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
QY 33 agcttcaaatccagaagatttcagcgtcttcaagaaggaagtataataatcctggc 92
DB 4360 AGGTTTAATAATATTTTCATCTTGATAGCATATGAATAATTATCTTTCATATAAGGCTT 4419
QY 93 ctggcagagaattgatgttcaatcgtatccatttgtaaaacagaatcacccatctg 152
DB 4420 CTTTAATAATAATAATATCTAGTGATCATTTTATGAATAATATTTTATGATTAT 4479
QY 153 ttcaatgcttaagaatgtttggaataatcttataataatcatttcagaggaattgact 212
DB 4480 ATTATATAAATAATATTTTATAATATCATATCACTAATTTTGAAAGATATTGTATA 4539
QY 213 tattgatgatgacccctataccccgaactatacagagga 255
DB 4540 ATTATATATTATTTTATAATAATAATTATATGCTTGTGTA 4582
RESULT 9
AC096291/c

LOCUS AC096291 116127 bp DNA linear HMG 20-DEC-2001
 DEFINITION Rattus norvegicus clone CH230-38P14, *** SEQUENCING IN PROGRESS
 ACCESSION AC096291
 VERSION AC096291.2 GI:17943988
 KEYWORDS HMG; HMGs_PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Rattus.
 1 (bases 1 to 116127)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alstbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbarella,J.,
 Benson,J., Bimagne,K., Blankenburg,K., Bonnin,D., Bouck,T.,
 Bowls,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
 Dem,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Ellhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flaagg,N., Ford,J.,
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
 Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
 Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hayes,A.,
 Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,
 Hollins,B., Homs,F., Howard,S., Huber,J., Hulys,S., Hume,J.,
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivet,S.,
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
 Louisedge,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
 Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
 Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabab,K.,
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkwo,S.,
 Ogutu,N., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
 Peery,J., Perez,L., Peters,L., Pickens,R., Prims,E., Pu,L.L.,
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojudoan,I., Rolfe,M.,
 Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshari,N.,
 Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,
 Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K.,
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,R.,
 Thomas,S., Usmani,K., Vasquez,L., Verra,V., Villalob,D., Vinson,R.,
 Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
 Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
 Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zortilla,S., Nelson,D.,
 Weinstein,G. and Gibbs,R.
 Direct Submission
 Unpublished
 2 (bases 1 to 116127)
 Worley,K.C.
 Direct Submission
 Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Dec 20, 2001 this sequence version replaced gi:15627911.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GEVI
 Center clone name: CH230-38P14
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329First call to
 findphraplist
 Consensus quality: 83123 bases at least Q40
 Consensus quality: 91227 bases at least Q30

Consensus quality: 96818 bases at least Q20
 Estimated insert size: 71455; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 0.9x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 61 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 1 3916: contig of 3916 bp in length
 3917 4016: gap of unknown length
 4017 7194: contig of 3178 bp in length
 7195 7294: gap of unknown length
 7295 11673: contig of 4379 bp in length
 11674 11773: gap of unknown length
 11774 14734: contig of 2961 bp in length
 14735 14834: gap of unknown length
 14835 18012: contig of 3178 bp in length
 18013 18112: gap of unknown length
 18113 19884: contig of 1772 bp in length
 19885 19984: gap of unknown length
 19985 22062: contig of 2078 bp in length
 22063 22162: gap of unknown length
 22163 25052: contig of 2850 bp in length
 25053 25152: gap of unknown length
 25153 27939: contig of 2787 bp in length
 27940 28039: gap of unknown length
 28040 30903: contig of 2864 bp in length
 30904 31003: gap of unknown length
 31004 33492: contig of 2469 bp in length
 33493 33592: gap of unknown length
 33593 36532: contig of 2940 bp in length
 36533 38863: gap of unknown length
 38864 38963: contig of 2231 bp in length
 38964 42247: gap of unknown length
 42248 42347: gap of unknown length
 42348 43940: contig of 1493 bp in length
 43941 47252: gap of unknown length
 47253 47352: contig of 3312 bp in length
 47353 50104: gap of unknown length
 50105 50204: contig of 2752 bp in length
 50205 52302: gap of unknown length
 52303 52402: contig of 2098 bp in length
 52403 54327: gap of unknown length
 54328 54427: contig of 1925 bp in length
 54428 55733: gap of unknown length
 55734 55833: gap of unknown length
 55834 57172: contig of 1339 bp in length
 57173 58476: gap of unknown length
 58477 58576: contig of 1204 bp in length
 58577 60321: gap of unknown length
 60322 60421: contig of 1745 bp in length
 60422 62141: gap of unknown length
 62142 62241: contig of 1720 bp in length
 62242 63722: gap of unknown length
 63723 63822: contig of 1481 bp in length
 63823 65055: gap of unknown length
 65056 65155: contig of 1233 bp in length
 65156 66343: gap of unknown length
 66344 66443: contig of 1188 bp in length
 66444 67996: gap of unknown length
 67997 69108: gap of unknown length
 69109 69208: contig of 1112 bp in length
 69209 69208: gap of unknown length

```
* 69209 70970: contig of 1762 bp in length
* 70971 71070: gap of unknown length
* 71071 72523: contig of 1453 bp in length
* 72524 72623: gap of unknown length
* 72624 74933: contig of 2330 bp in length
* 74934 75053: gap of unknown length
* 75054 76089: contig of 1036 bp in length
* 76090 76189: gap of unknown length
* 76190 77687: contig of 1498 bp in length
* 77688 79293: gap of unknown length
* 79294 79393: gap of unknown length
* 79394 80715: contig of 1322 bp in length
* 80716 82192: gap of unknown length
* 82193 82292: gap of unknown length
* 82293 83772: contig of 1480 bp in length
* 83773 83872: gap of unknown length
* 83873 84994: contig of 1122 bp in length
* 84995 85094: gap of unknown length
* 85095 86566: contig of 1562 bp in length
* 86567 86756: gap of unknown length
* 86757 88212: contig of 1456 bp in length
* 88213 89343: gap of unknown length
* 89344 89443: gap of unknown length
* 89444 90867: contig of 1424 bp in length
* 90868 90967: gap of unknown length
* 90968 92392: contig of 1425 bp in length
* 92393 92492: gap of unknown length
* 92493 94094: contig of 1602 bp in length
* 94095 94194: gap of unknown length
* 94195 95340: contig of 1146 bp in length
* 95341 95440: gap of unknown length
* 95441 97148: contig of 1708 bp in length
* 97149 97248: gap of unknown length
* 97249 98441: contig of 1193 bp in length
* 98442 98541: gap of unknown length
* 98542 99693: contig of 1152 bp in length
* 99694 99793: gap of unknown length
* 99794 101143: contig of 1350 bp in length
* 101144 102563: contig of 1320 bp in length
* 102564 102663: gap of unknown length
* 102664 103974: contig of 1311 bp in length
* 103975 104074: gap of unknown length
* 104075 105257: contig of 1183 bp in length
* 105258 105357: gap of unknown length
* 105358 106636: contig of 1279 bp in length
* 106637 106736: gap of unknown length
* 106737 107801: contig of 1065 bp in length
* 107802 107901: gap of unknown length
* 107902 109084: contig of 1183 bp in length
* 109085 109184: gap of unknown length
```

Query Match 8.4%; Score 42; DB 2; Length 116127;
Best Local Similarity 50.0%; Pred. No. 7.6; 105; Indels 0; Gaps 0;
Matches 105; Conservative 0; Mismatches 105;

```
QY 255 aagtcagacatgtgtgaacatcattgacatgcagatcattcatgcgttatcatttagagc 314
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 75355 AAGATCATTAAGTATGTGCGCAAAATAGAAATTATAGAGTTTATATTAAGATGGACA 75296
QY 315 atgaatctgaagcagcctccagaaatttttagcagacattgttacaanaacttgaggggt 374
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 75295 TTAAAAATGATTAACATTAATCCATTTTAAGAATTCATPAGAAAGAAAAATGCAATTG 75236
QY 375 ttgtccttaaaaatttagatgcatctatacaaacatccgccttataaaaatatattcatt 434
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 75235 TGAATTAATAGTTTAATTAATTAATCAACTACTATTATTCAAAAATAAAGTTTACGACAGA 75176
QY 435 caccattaccaagtcgtgttttgcagataa 464
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 75175 FACAGAAAGAAATTTCTGACTTCACATATA 75146

RESULT 10
CEY39A1A 110960 bp DNA linear INV 11-DEC-2001
LOCUS Caenorhabditis elegans cosmid Y39A1A, complete sequence.
DEFINITION AL031633.299293
ACCESSION AL031633.1 GI:3646884
VERSION HTG.
KEYWORDS Caenorhabditis elegans.
SOURCE Caenorhabditis elegans.
ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (sites)
AUTHORS none.
TITLE Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium
JOURNAL Science 282 (5396), 2012-2018 (1998)
MEDLINE 99069613
REMARK The C.elegans Sequencing Consortium.
REFERENCE 2 (bases 1 to 110960)
AUTHORS Wall,M.
TITLE Direct Submission
JOURNAL Submitted (22-SEP-1998) Nematode Sequencing Project, Sanger
Institute, Hinxton, Cambridge CB10 1SA, England and Department of
Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rwnematode.wustl.edu
Coding sequences below are predicted from computer analysis, using
predictions from GeneFINDER (P. Green, U. Washington), and other
available information.
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is not the entire insert of clone Y39A1A.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone Y39A1 is at 1 in this sequence. The true
left end of clone M09D10 is at 110854 in this sequence. The start
of this sequence (1..106) overlaps with the end of sequence
AL032621.
The end of this sequence (110854..110960) overlaps with the start
of sequence Z93785.
For a graphical representation of this sequence and its analysis
see: [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=Y39A1A)
[name=Y39A1A](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=Y39A1A)
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
Location/Qualifiers
1..110960
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="III"
/clone="Y39A1A"
complement(join(1078..1247,2168..2941,3412..3482,
3561..3777,4206..4296))
/gene="Y39A1A.1"
complement(join(1078..1247,2168..2941,3412..3482,
3561..3777,4206..4296))
/gene="Y39A1A.1"
/note="cDNA EST EMBL:T00867 comes from this gene"
/codon_start=1
/protein_id="CAA21019.2"
/db_xref="GI:15718274"
/db_xref="SPTREMBL:O9X921"
/translation="MSKKEEITFEFRLENRPRESSKKEEDENSTEEEMTLNHNASYTLD
HSAPFIADKDFKYNOLNPLHFRMYKDYKGGSSRIYENIFLSKLYKIFCGILK
QGVLTAKINFTYHNFTSTWPLVKIKNGPVLVLYVDGNSRRRIIYVSALAGRPQNN

```

gene
CDS
    join(14209..14677,14977..15126,15303..15440,15752..16004,
    16352..16433,16754..16939,18701..19006)
    /gene="Y39A1A.5a"
    join(14209..14677,14977..15126,15303..15440,15752..16004,
    16352..16433,16754..16939,18701..19006)
    /gene="Y39A1A.5a"
    /note="contains similarity to pfam domain: PR02204
    (Vacuolar sorting protein 9 (VPS9) domain), Score=36.7,
    E-value=1.7e-07, N=1"
    cDNA EST yk210c11.3 comes from this gene
    cDNA EST yk210c11.5 comes from this gene
    cDNA EST yk532c5.5 comes from this gene
    cDNA EST yk556b4.5 comes from this gene"
    /codon_start=1
    /protein_id="CAA21021.1"
    /db_xref="GI:3880848"
    /db_xref="SPTREMBL:Q9XX19"
    /translation="MAQETARSVKNAVKISQEMRMSDMSSELYWSYVYIGRTGCH
    SLFSPDCKVKEVDYEDVQVEKYISFECSIFPCANHEEYVADMSQIDIRSLHWYTAG
    FLETRKWKQOTRDKIDEAISLEIENAKSAFELDLCTKSCAITEALKESSNI
    ISYKPGLYWYMKWKMLKKNFNOKIRKTSKIRFEVFKMYFLTRKASTSADFLPT
    LIYVLRGNPLQISNVKFIISRFALPARLMSGEAYFTNLSCAFGRNNHESLQK
    KSESEAYTSGHLAPLSVINSACNOALIVYEGETITTNKAGKSLAKNLNHHGK
    SDDLEKMLAVIKETVDYPTDEYMNKMSIPEAEKETADILVLSRSSSSGSRRTS
    DPOSTROPDSTISPLAETAOINLTSSGNOBHGEGSLAAGPILLEMALVYKIAS
    IFKQAGAGNTVSASKRGSTQSPVITRASTORASPAIAQVEKESVQKPREPSRSK
    HHAHRSKMSALKAASON"
    join(18036..18110,18701..19006)
    /gene="Y39A1A.5b"
    join(18036..18110,18701..19006)
    /gene="Y39A1A.5b"
    /note="cDNA EST yk112f11.3 comes from this gene
    cDNA EST yk112f11.5 comes from this gene
    cDNA EST yk68h1.3 comes from this gene
    cDNA EST yk68h1.5 comes from this gene
    cDNA EST yk451e4.3 comes from this gene
    cDNA EST yk437d12.3 comes from this gene
    cDNA EST yk437d12.5 comes from this gene
    cDNA EST yk532c5.3 comes from this gene
    cDNA EST yk540g3.3 comes from this gene"

```

```

gene
CDS
    join(19588..19881,20463..20639,20856..20982,21728..21912)
    /gene="Y39A1A.6"
    join(19588..19881,20463..20639,20856..20982,21728..21912)
    /gene="Y39A1A.6"
    /note="cDNA EST yk471b2.3 comes from this gene
    cDNA EST yk471b2.5 comes from this gene"
    /codon_start=1
    /protein_id="CAA21022.1"
    /db_xref="GI:3880849"
    /db_xref="SPTREMBL:Q9XX20"
    /translation="MLRLVLRRTGSEACAAVSTRAVLVSHPVYVSEDSLSAKETDRRS
    KLRYPKLKDQSLSKVYFAPEWLEKRPNLDEGVNPLKGYMPEKWEYKRVKVP
    PNYIVPEGLPKPEVEFKESVHSPKRMQAQOLVWKMVDAITQDMQOLKACN
    LLMPTIRKAKSRADDEHIEYPSQMYVADAPVOSNIVKGRAHADMWMTIRYRIH
    IFVRLSEGPAPQOORHPQKNGMDHDEYVYVLSRYKYSI"
    complement(join(22174..22353,23082..23416,24281..24329))
    /gene="Y39A1A.7"
    complement(join(22174..22353,23082..23416,24281..24329))
    /gene="Y39A1A.7"
    /note="predicted using GeneFinder"
    /codon_start=1
    /protein_id="CAA21015.1"
    /db_xref="GI:3880842"
    /db_xref="SPTREMBL:Q9XX25"
    /translation="WTSITSAFLILLFVAVACNSVPLDCTCELSHDPVIRCDGV
    GINRPLPSSHSPLRGFHFLATCDIDITIPASIMKASFPLDQIDVQGNRLKNCSDL
    VHLHNDIVLDCDCKNEKPLQCDNLDKCDMKCRFLTKLEKMAQFQDLVNRKAEWQA
    BETTEAKSWFSAQOFKFKRIAGELSD"
    join(25750..25789,25839..25909,25963..26088,26137..26405,
    27339..27744,28541..28596,28650..28837,28993..29064,
    29500..29707,30169..30268,30345..30355)
    /gene="Y39A1A.23"
    join(25750..25789,25839..25909,25963..26088,26137..26405,
    27339..27744,28541..28596,28650..28837,28993..29064,
    29500..29707,30169..30268,30345..30355)
    /gene="Y39A1A.23"
    /note="cDNA EST EMBL:AU108961 comes from this gene
    cDNA EST EMBL:AU112975 comes from this gene"
    /codon_start=1
    /protein_id="CAC42372.1"
    /db_xref="GI:14530622"
    /translation="WKGQVAVQSNLKTMSRSIALSKISDVLLVSEGLFRTYVN
    RSKCVFPEFAPEFNACDVSMINKKAVICLSMKSAIRIKGVAFGKKNVGEFRI
    DPKAEKRMVQLOMNYKGIISCRILEKLAIEDIRITIAKREMGSMHKGVEKARSG
    RNITVYFASLTLPFVQKGDIEVTKMVDGGLDIRNSHSDGVTMFMGVEKARSKY
    KTEPTITQEKTRKIOIVPERSISIKELTSLVFPADQSGEYVMYDLPEKPLIVSI
    EAHNRFDIELMAGSDDELDDOGGILKETMAQHEEEDSTAHSSSRKSKAIDP
    SSSGTRKSKCESISLQETTRTOSLPSRKNFPELIPVADQSMRDREYVTEQREPS"

```

Query Match 8.3%; Score 41.6; DB 3; Length 110960;
 Best Local Similarity 48.7%; Pred. No. 9.4;
 Matches 113; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

```

QY 165 agatggttggaanaatttatatacatcttcagagagattgacttatttgatgaga 224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 17594 ACAAAAATTAATAAACTTCAACAACTAAATTTGAAAATTCGAAATTTTTCGATTTT 17535
QY 225 tgcacattatcccgagaactatcagagagcgcagacatgcttggaaccatgac 284
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 17534 CCAGGTTTTTTTCCCAACATTTTTCACAAATGAGATTAAATTTTTCGATTTTTC 17475
QY 285 tcagattccatcgttagcatttagaagatcgaatcgaagcacagcttcagaaattt 344

```


Db 17474 CCAAAACATCAATATAGTTTTCATTTGGAAATTTGGCATTTT 17415
 QY 345 tagcagacattgtacaaacctggagggttcttccaaataatgatgc 396
 Db 17414 TTCTAGAAATTTTAAATAATTTTTCGAATTAATAATCCGCTAC 17363

RESULT 11

AC060835 160624 bp DNA linear PRI 09-JAN-2002
 LOCUS Homo sapiens BAC clone RP11-785F11 from 4, complete sequence.
 AC060835
 AC060835.7 GI:15145631
 HTG.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 160624)

REFERENCE Sulston, J.E. and Waterston, R.
 TITLE Toward a complete human genome sequence
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
 MEDLINE 99063792

REFERENCE Tomlinson, C., Abbott, A., Hawkins, M., Dixon, R. and Boyer, E.
 TITLE The sequence of Homo sapiens BAC clone RP11-785F11
 JOURNAL Unpublished (2001)
 3 (bases 1 to 160624)

REFERENCE Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (20-APR-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

REFERENCE 4 (bases 1 to 160624)
 WATERSTON, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (09-AUG-2001) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

REFERENCE 5 (bases 1 to 160624)
 WATERSTON, R.
 TITLE Direct Submission
 JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Aug 9, 2001 this sequence version replaced g1:14670156.

COMMENT

Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: saplensew@wustl.edu
 Summary Statistics
 Center project name: H_NH0785F11

NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping
 clone sections once, or longer because we provide a small overlap
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate
 chemistry, or covered by high quality data (i.e., phred quality >= 30);
 an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by sequence
 from more than one subclone; and the assembly was confirmed by
 restriction digest.

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. John D.
 McPherson, Department of Genetics, Washington University, St. Louis
 MO. For additional information about the map position of this
 sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male
 donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,
 Tatem, M., Catene, J.J. and de Jong, P.J. (1998) An improved
 approach for construction of bacterial artificial chromosome
 libraries. Genomics 51:1-8. The clone may be obtained either from
 Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong
 and coworkers at the Roswell Park Cancer Institute
 (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is RP11-114M4; the clone sequenced
 to the right is RP11-327J3. Actual start of this clone is at base
 position 1 of RP11-785F11; actual end is at base position 160624 of
 RP11-785F11.

FEATURES

source Location/Qualifiers
 1..160624
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="4"
 /map="4"
 /clone="RP11-785F11"
 /clone_lib="RP11-11"
 78..478
 /rpt_family="L2"
 497..1140
 /rpt_family="L1"
 3053..3103
 /rpt_family="AT_rich"
 3390..3638
 /rpt_family="L2"
 3656..4189
 /rpt_family="ERV_L"
 4201..4975
 /rpt_family="L2"
 4983..5188
 /rpt_family="MER1_type"
 5193..5243
 /rpt_family="L2"
 5299..5534
 /rpt_family="L2"
 6010..6383
 /rpt_family="MALR"
 6471..7561
 /rpt_family="L1"
 7558..8936
 /rpt_family="L1"
 8937..9234
 /rpt_family="Alu"
 9235..9420
 /rpt_family="L1"
 9388..9484
 /rpt_family="L1"
 9489..10100
 /rpt_family="L1"
 10114..11038
 /rpt_family="L1"
 11039..11289
 /rpt_family="Alu"
 11290..13382
 /rpt_family="L1"
 13375..13395
 /rpt_family="AT_rich"
 13396..13440
 /rpt_family="(TA)n"
 14049..14187
 /rpt_family="MER1_type"
 15199..15223
 /rpt_family="(A)n"
 16128..16439
 /rpt_family="L1"
 16440..16743
 repeat_region

```
repeat_region /rpt_family="Alu" 16744. 16874
/rpt_family="L1"
repeat_region /rpt_family="ERV1" 16916. 17257
/rpt_family="ERV1" 17413. 17440
repeat_region /rpt_family="AT-rich" 18356. 18659
/rpt_family="Alu" 18631. 18671
repeat_region /rpt_family="(CAAA)n" 18807. 18859
/rpt_family="AT-rich" 18884. 19097
/rpt_family="L1" 19070. 19091
repeat_region /rpt_family="AT-rich" 19110. 19547
/rpt_family="L1"
repeat_region /rpt_family="T)n" 21198. 21237
/rpt_family="(T)n" 21214. 21497
/rpt_family="Alu" 23018. 23045
repeat_region /rpt_family="AT-rich" 23028. 23698
/rpt_family="L1" 23692. 24220
repeat_region /rpt_family="L1" 24308. 25255
/rpt_family="L1" 30504. 31045
/rpt_family="L1" 31156. 31254
repeat_region /rpt_family="MIR" 31564. 31595
/rpt_family="AT-rich" 31794. 31882
repeat_region /rpt_family="AT-rich" 32101. 32426
/rpt_family="ERV1" 32394. 32414
/rpt_family="AT-rich" 32418. 32479
repeat_region /rpt_family="(TC)n" 32488. 32523
/rpt_family="(CA)n" 32613. 32667
repeat_region /rpt_family="(CAGAGA)n" 33504. 33524
/rpt_family="AT-rich" 34257. 34475
repeat_region /rpt_family="MIR" 34748. 35031
/rpt_family="Alu" 35304. 35454
repeat_region /rpt_family="MIR" 35602. 35667
/rpt_family="ERV1" 37934. 39378
repeat_region /rpt_family="L2" 38783. 38810
/rpt_family="(CA)n" 42850. 42890
repeat_region
```

OY	183	tctatatacatttttcagaggatttgacttatttgtgatgaagtgcacctgtccoccgaa	242
Dn	19221	TGGATGTGCATTTCAGAAAAAATAAAAGAATTGTTACCTCCTCCACAATACCATTATCGTGAAT	19280
OY	243	cctacaccgagaagcgtcacgacatgctggtagcaattacagactcagaattcccatcgtag	302
Dn	19281	TAAITTCAGATGCGATTTTAGACTTAATATGTCAAGAGATNAGAACACAGAGTGCTATNAGA	19340
OY	303	tcatttagacgaatgaatcgtgagcacacacgccctccagaatlitttagcagacaattgbtaaaa	362
Dn	19341	AACATTGTGACATTATTTTTTAAGTAGTTCTTAAACTGTGATTCAAAAGCACTAACAATA	19400
RESULT 12			
LOCUS	AC020669/c		
DEFINITION	Homo sapiens clone RP11-114M4, WORKING DRAFT SEQUENCE, 5 unordered pieces.		
ACCESSION	AC020669	168982 bp DNA linear HTG 26-MAY-2000	
VERSION	AC020669.4 GI:7107781		
KEYWORDS	HTG; HTGS_PHASEL; HTGS_DRAFT.		
SOURCE	human.		
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 		
REFERENCE	Bliren,B., Linton,L., Nusbaum,C. and Landier,E. Homo sapiens, clone RPl1-114M4 Unpublished 2 (bases 1 to 168982)		
AUTHORS	Anderson,S., Baldwin,J., Barra,N., Beckertly,R., Bede,F., Boguslavsky,I., Boukhalter,B., Brown,A., Burkett,G., Castle,A., Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cooke,P., Darellano,K., Dewar,K., Domino,M., Doyle,M., Fensteron,J., Ferreira,P., FitzHugh W., Forrest,C., Gage,D., Galagan,J., Gerdyana,S., Grant,G., Hagos,B., Heatford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Kartas,A., Klein,J., Landers,T., Lehoczyk,J., Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marguis,N., McNameis,P., McGurk,A., McKernan,K., McNeethers,R., Meldrum,J., Menaus,L., Morrow.J., Naylor,J., Norman,C.H., O'Connor.T., O'Donnell.P., Olivar,T.M., Peterson.K., Pierre,N., Pisanl,C., Pollard,V., Raymond,C., Riley.R., Kochman.D., Roy.A., Santos,R., Severy,P., Spencer.B., Stange-Thomann.N., Schojanovic,N., Subramanian,A., Talamas,J., Tesfaye.S., Theodore,J., Tirrell.A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman.D., Ye,W.Y., Zimmer.A. and Zody M.		
TITLE	Direct Submission		
JOURNAL	Submitted (08-JUN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA On Feb 28, 2000 this sequence version replaced gi:6751804. All repeats were identified using RepeatMasker: Shult, A.F.A & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html		
COMMENT	----- Genome Center Center : Whitehead Institute/ MIT Center for Genome Research Center code : WIBR Web site: http://www-seq.wi.mit.edu Contact: sequence_submissions@genome.wi.mit.edu ----- Project Information Center project name: LI948 Center clone name: LI4_M_4 ----- Summary Statistics Sequencing vector: MI3: M7815; 100% of reads Chemistry: Dye-terminator Big Dye; 100% of reads Assembly program: Phrap; Version 0.960731 Consensus quality: 167391 bases at least Q40 Consensus quality: 168285 bases at least Q30 Consensus quality: 168505 bases at least Q20 Insert size: 168000; agarose-fp Insert size: 168582; sum-of-configs Quality coverage: 6.0 in Q20 bases; sum-of-configs Quality coverage: 6.0 in Q20 bases; sum-of-configs		

Db 30174 TGGATATGCAATTGAGAAAAAAGTAATTTAGCTCTCCACATACCATATCCGTGAAT 30233
 Qy 243 ctactactagaaggaagtcacacatgltgtggaactcatcagatccagatccatcgttag 302
 Db 30234 TAATTCAGATGCAATTTTACACTAATATGGAAGATAGACACACAGCTCTTAGAGA 30293
 Qy 303 tcaatttagcagatgaatcgtgaacacagctccacagaatttttagcagatcgttagcaaa 362
 Db 30294 AACAATTCGAGATTATTTATTAAGTAGTCTCTTAACCTAGATTCAAAAGCATTAACATA 30353

RESULT 14
 AC087232/c 49623 bp DNA linear INV 18-OCT-2001
 LOCUS Caenorhabditis elegans cosmid Y92H12BR, complete sequence.
 DEFINITION AC087232
 AC087232
 AC087232.2 GI:12863250
 KEYWORDS HTG.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 49623)
 The C. elegans Sequencing Consortium.
 Genome sequence of the nematode C. elegans: a platform for
 investigating biology. The C. elegans Sequencing Consortium
 Science 282 (5396), 2012-2018 (1998)
 9069613
 2 (bases 1 to 49623)
 Tin-Wollam, A., Wollmann, P. and Courtney, L.
 The sequence of C. elegans cosmid Y92H12BR
 Unpublished
 3 (bases 1 to 49623)
 Waterston, R.
 Direct Submission
 Unpublished
 4 (bases 1 to 49623)
 Waterston, R.
 Direct Submission
 Submitted (19-DEC-2000) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 5 (bases 1 to 49623)
 Waterston, R.
 Direct Submission
 Submitted (16-FEB-2001) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 6 (bases 1 to 49623)
 Waterston, R.
 Direct Submission
 Submitted (23-JUN-2001) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 7 (bases 1 to 49623)
 Waterston, R.
 Direct Submission
 Submitted (18-OCT-2001) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 On Feb 16, 2001 this sequence version replaced gi:11890811.
 Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1RO, England
 email: rtw@emataedu.wustl.edu and jess@sanger.ac.uk

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

This sequence was finished as follows unless otherwise noted: all

regions were double stranded, sequenced with an alternate chemistry
 or covered by high quality data (i.e., phred quality >= 30); an
 attempt was made to resolve all sequencing problems, such as
 compressions and repeats; all regions were covered by sequence from
 more than one ml3 subclone.

NOTES:

Coding sequences below are predicted from computer analysis, using
 the program GeneFinder (P. Green and L. Hillier, ms in preparation).

FEATURES

source
 1..49623
 /organism="Caenorhabditis elegans"
 /strain="Bristol N2"
 /db_xref="taxon:6239"
 /chromosome="I"
 /clone="Y92H12BR"
 2892..11265
 /gene="Y92H12BR.7"
 join(2892..3103,4082..4220,4262..4484,5839..6030,
 7491..7670,10415..11265)
 /gene="Y92H12BR.7"
 /codon_start=1
 /product="Hypothetical protein Y92H12BR.7"
 /protein_id="AAK09079.1"
 /db_xref="GI:12863254"
 /translation="MSPNGMQPMPHNOASRPDSMPVANNNGGPIROPPOAORS
 NCMIVLANPYFSRVGAQSHYTYERNRMVFSPOPYVQKPLISQAPPPNPSEF
 VQTLMSKRIPEPPPTPMFLDSVRVSKCMLSNQKLEKRRERQDMQFOSNFOF
 EPRKLSFDDSSRQREVCAGFAFLFWSPDHDIDIFLTSTHTTKTKMTSEDPQSS
 KNYELQVTAAGDITAEIKCISTDKDFLVAIWSERAEFSYSELRLQKCYOFQRG
 IPNAEPLMEHFCVVRCKRYWDEIHHCNCNVOMOKYIKDNGCGPGLNGKINGLFLF
 SARLLPDLITLPOFSPGPNRMVINGFVILNPERHNFYADFCNKNHYVMYVGLVN
 SETDIYCREKLYKIMPLSNPEYKVLPPYVDQGRFVYNTLWLYETDIQIDMG
 QPSALMATGAGSYRIGLGPNNQCTMCNLYPIGKKLLEPAVPEVASVKNVDSMTTII
 LREGGHDSEVADTLAYLIDREERSLTAQELAKRLDELVKAVDHNKNAVAREKNA
 SMTSIFRNLDNDFGFRKRDALFEIIRKLKN"
 14267..21355
 /gene="Y92H12BR.6"
 join(14267..14887,20972..21355)
 /gene="Y92H12BR.6"
 /note="coded for by the following C. elegans cDNAs:
 yk258f5.5, yk372d4.5, yk258f5.3"
 /codon_start=1
 /product="Hypothetical protein Y92H12BR.6"
 /protein_id="AAK09076.2"
 /db_xref="GI:16259269"
 /translation="MNEFISWCKGRGYQFDGLEITCPPPGNGGNGIVATGFRGKAF
 ITLPEEDMINALVVDLPIYRKKLAKIGGEKLPMEILTMFEAFEDTSHSAMPYLV
 LPKENTPAFGIDVDNMLPLSIRKYMDOKKEISEKVKSVTKNPIFLRSGV
 FLOKSKKFRKIAKNFNSPIFMFTQRAKREKRIASNSNFRFEMWTREHVELEE
 VGLPSHLYATDGGPSWSLRTNIRVNLLEHWMANRKKIITISHPDIDDEKIDEX
 LEIMLEKLNQVAERAKKMPIDVKMMWEDQVACVVALQSEFKKLEKEKLEEYS
 "
 complement(22391..24116)
 /gene="Y92H12BR.2"
 complement(join(22391..22562,23561..23672,23723..23810,
 23876..23969,24049..24116))
 /gene="Y92H12BR.2"
 /note="coded for by the following C. elegans cDNAs:
 yk114a10.3, yk114a10.5, yk254e10.3, yk254e10.5, yk337h7.3,
 yk337h7.5, yk605b11.3, yk605b11.5"
 /codon_start=1
 /product="Hypothetical protein Y92H12BR.2"
 /protein_id="AAK09077.1"
 /db_xref="GI:12863252"
 /translation="MTTKRVPSKIVIASSKITLNDPESAIIPKPVKATVMAKQPKRF
 SAQPKRFSVEEVEYVPEKMKPRKPFQIQOARVITQNRVSVNMAVPQRFVPR
 NNEVRSNOMQNEPKPRPQNFHNRKMFQGNRRRPHRQKQPTPAKKSVAELD
 RELDDYMRKPKAPITIT"
 24914..32931
 /gene="Y92H12BR.3"
 join(24914..25060,25118..25283,28534..29631,31011..31315,

gene
 CDS

```

32734..32931)
/gene="y92H12BR.3"
/notes="coded for by the following C. elegans cDNAs:
YK417612.5, YK52597.5, YK5616.5, YK609H10.5, YK644F4.5,
YK644F4.3, YK609H10.3, YK52597.3"
/codon_start=1
/product="Hypothetical protein Y92H12BR.3"
/protein_id="AK09080.2"
/db_xref="GI:16259268"
/translation="MKAQLALISLWGLLAADLRSLIDICAMPKPTTKTSDKIST
PESEKISGSMTPDMKASTLSLVSASKEPRLVKQRDMQPEKWIQELTAKLTSEFN
VSGEKTGNVSKPDLKISRESEIIGADTSLSLILGLIEPKSHGFEVEHEVY
GGVNAQKWTICIDGATNDPKVLEVRVAGEETIRASQSPFSNPLDISLAEIPN
STVALNHVSVELDRYELPYGNEAAVHVAGCKNRNDIVLKLDEYAAILNEDPST
NRDVEVLYSKTKVYIHAGDSFENGVKLFKATPKRYKGTQYILHDFNYGEFTS
HGACQSEPLDSEDEVLYKNGMDHFRIMNMEDLIVDPCISLSESDVMQTAFT
YRAYDSRDETVELIHTEDMEVHSMARFRNGSHLAVSLVQITPMESSLNKKTVOL
SNQIDQKFRANNTLAVQIKDKTSLDSQVGLNLENAVASSIKNVHPIIPRYVY
VESRDSGLEIVAVAKTERKPGNVVNTTELTALTEIFKLMNDSKTDLKMPIEVLK
NNGVEEMMADASTMKMYPENDGFIYTGAMFVLAIFCLIGVIGSAGVAVATR
RORISTLAYOVFE"
/complement(41338..42018)
/gene="y92H12BR.4"
/complement(join(41338..41475,41971..42018))
/notes="coded for by the following C. elegans cDNAs:
YK43295.3, YK43295.5"
/codon_start=1
/product="Hypothetical protein Y92H12BR.4"
/protein_id="AK09081.1"
/db_xref="GI:12863256"
/translation="MDTSPTSCAGVFQNLDNLSIQVENNQIIVYNILENVGDIIND
RLKLRVSHFDNHCIFL"
/4615..49537
/gene="y92H12BR.8"
/join(4615..44752,45702..45826,48904..49537)
/gene="y92H12BR.8"
/notes="coded for by the following C. elegans cDNAs:
YK25796.5, YK344F6.5, YK440F7.3, YK440F7.5, YK454F2.5,
YK601D9.5"
/codon_start=1
/product="Hypothetical protein Y92H12BR.8"
/protein_id="AK09082.1"
/db_xref="GI:12863257"
/translation="MSQARSKERALQIYEASRIRIODARNIGARTGROLKKGH
NOMGHTOGALERAKPPLGIMNGDFRPMQRMVPGKELFNADINTREYIPLSIVELA
RLTLDLMINPROPIDISTCTOKFOINPMROYGDELTEGADSEFYSIDIEVQAT
PSAIAAVERAGGVRRTAYVDVLEAIAINKCFEKGKVPKRPQPSILLEYMDAK
NRQYLCEDEIEKERALSAQIRNGYTLPPPPSSAALKAVDQVFGIIPSGSVSLDRKY
FVTKNELHREYSSRADKLKYS"
BASE COUNT 16759 a 8291 c 7983 g 16590 t
ORIGIN
Query Match 8.3%; Score 41.4; DB 3; Length 49623;
Best Local Similarity 46.6%; Pred. No. 12;
Matches 132; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

```

```

QY 149 tcggttcaggcctaagaaggttggaaaaaattctatacaatttccagaagattg 208
DB 45600 TTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 45541
QY 209 acttattgatgagatgcacactatccacagaactactagaggagagcagacagt 268
DB 45540 AATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 45481
QY 269 gtggaactactagactcagaatcactcgttagctactttaagaagatgaactgaagca 328
DB 45480 TTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 45421
QY 329 cagcttcagaatcttaagcagacatgtaaaaaaacttggagggttgcctctaaaaaa 388
DB 45420 TTTTCTACTAAATTTTGAATTTTGGCAAAAATTTGTTTAAATTTTCCGTTAATTTT 45361

RESULT 15
CBRG01C7/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA
e-mail: jsplet@watson.wustl.edu

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.
Location/Qualifiers
1..39419
/organism="Caenorhabditis briggsae"
/strain="Gujarat G16"
/db_xref="taxon:6238"
/clone="G01C7"
BASE COUNT 12588 a 7352 c 7744 g 11735 t
ORIGIN
Query Match 8.0%; Score 40.2; DB 3; Length 39419;
Best Local Similarity 45.2%; Pred. No. 22;
Matches 147; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

```

Db 35243 ACAATCCAGATAGTGCATCAAGAA 35219

Search completed: May 22, 2002, 06:36:01
Job time: 7746 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 04:29:50 ; Search time 463.88 Seconds
(without alignments)
1850.602 Million cell updates/sec

Title: US-09-693-205-7_COPY_1_500
Perfect score: 500
Sequence: 1 atgatttcaggaagacatc.....gcagaattgttaatcaaa 500

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq_032802.*
1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	500	100.0	12792	22	AAH20176
2	500	100.0	12793	22	AAH20174
3	500	100.0	12793	22	AAH20178
4	500	100.0	12793	22	AAH20179
5	500	100.0	12793	22	AAH20182
6	331.2	66.2	11493	22	AAH20175
7	42.2	8.4	1251	21	AAH70119
8	38.4	7.7	1632	22	AAH32525
9	38.4	7.7	1892	22	AAH02544

C	10	38.4	7.7	3644	23	ABL12784	Drosophila melanog
C	11	37.8	7.6	513445	22	AA161373	Soybean 318013 reg
C	12	37.4	7.5	7957	23	ABL26428	Drosophila melanog
C	13	37.4	7.5	11172	24	ABL34053	Human immune syste
C	14	37.2	7.4	3129	23	AA555610	Streptococcus pneu
C	15	37.2	7.4	39003	22	AA28534	Genomic fragment #
C	16	37.2	7.4	40862	22	ABL34073	Human immune syste
C	17	36	7.2	11410	20	AA131327	Enterococcus faeca
C	18	36	7.2	1038602	20	AA201425	Complete genome se
C	19	35.6	7.1	516	20	AA39579	Nucleic acid sequ
C	20	35.6	7.1	2531	19	AA42968	Streptococcus pneu
C	21	35.6	7.1	3102	22	AAH90796	CYE 100 coding seq
C	22	35.6	7.1	3129	22	AA691296	Streptococcus pneu
C	23	35.6	7.1	19446	19	AAV52184	Streptococcus pneu
C	24	35.4	7.1	735	22	AA196604	Human neuroblastom
C	25	35.4	7.1	1664976	19	AAV21209	Methanococcus jann
C	26	35.2	7.0	1556	22	AA05528	Human secreted pro
C	27	35.2	7.0	2325	22	AAH15284	Human cDNA sequenc
C	28	35.2	7.0	2895	22	AA05496	Human secreted pro
C	29	35	7.0	965	20	AA134459	Enterococcus faeca
C	30	35	7.0	6536	24	ABL32147	Human immune syste
C	31	35	7.0	8378	22	AAH55763	Human adult form o
C	32	35	7.0	8378	22	AAH55764	Human neonatal for
C	33	35	7.0	32185	22	AA534542	Human DNA for a no
C	34	34.8	7.0	427	22	AA125826	Human breast cance
C	35	34.8	7.0	505	22	AA18992	Human breast cance
C	36	34.8	7.0	544	23	AA552216	Staphylococcus aur
C	37	34.8	7.0	546	23	AA54963	Staphylococcus aur
C	38	34.8	7.0	658	21	AA280691	Human colon cancer
C	39	34.8	7.0	744	22	AAK91766	Human cDNA 5'-end
C	40	34.8	7.0	744	22	AAK93599	Human cDNA clone r
C	41	34.8	7.0	842	18	AAV75119	Staphylococcus aur
C	42	34.8	7.0	1323	22	AAH72783	Human cervical can
C	43	34.8	7.0	1635	22	AAH94210	Human full-length
C	44	34.8	7.0	1711	22	AAH75817	Human G protein 44
C	45	34.8	7.0	1724	21	AAH16694	Human secreted pro

ALIGNMENTS

RESULT	1	
AAH20176		
ID	AAH20176	standard; DNA; 12792 BP.
AC	AAH20176;	
XX		
DT	09-AUG-2001	(first entry)
XX		
DE		Human mutated spastin nucleotide sequence SEQ ID NO:7.
XX		
KW		Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
KW		autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KW		neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KW		reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
KW		atrophy of upper cerebellar vermis; absence of purkinje cell;
KW		abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
XX		
OS		Homo sapiens.
XX		
OS		Synthetic.
XX		
FH		Location/Qualifiers
FT		/*tag= a
FT		77..6604
FT		/product= "mutated spastin"
XX		
PN		WO200129266-A2.
XX		
PD		26-APR-2001.
XX		
PF		20-OCT-2000; 2000WO-US29130.
XX		
PR		20-OCT-1999; 99US-0160588.

XX (UYMC-) UNIV MCGILL.
 PA (HOP1-) HOPITAL SAINTE-JUSTINE.
 XX
 PI Hudson TJ, Engert J, Richter A;
 XX WPI: 2001-308494/32.
 DR P-PSDB: AAB97821.
 XX
 PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -
 XX
 PS
 XX
 PS Claim 1: Page -: 76pp: English.

CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSA) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with spastin
 CC protein or spastin gene expression or activity. The present sequence
 CC encodes a mutated human spastin.
 CC N.B. The present sequence is not given in the present specification but
 CC is derived from the human spastin nucleotide sequence (AAH20174) as
 CC stated on page 14.
 CC
 XX
 SQ Sequence 12792 BP: 4163 A; 2256 C; 2487 G; 3886 T; 0 other;

Query Match 100.0%; Score 500; DB 22; Length 12792;
 Best Local Similarity 100.0%; Pred. No. 8,6e-127;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgattacagagagacatgactacgctcagctccttaataccagaagattgacg 60
 DB 1 atgattacagagagacatgactacgctcagctccttaataccagaagattgacg 60
 QY 61 tcttatacagaagatgatacatctgctgctgcagagaattgatttcaatgta 120
 DB 61 tcttatacagaagatgatacatctgctgctgcagagaattgatttcaatgta 120
 QY 121 tccatttgatgaagaacagaatacccatctgtttcatagcttaagaatgttggaanaa 180
 DB 121 tccatttgatgaagaacagaatacccatctgtttcatagcttaagaatgttggaanaa 180
 QY 181 tcttatacatttccagagattgacttatttgatgagatgcacttaccag 240
 DB 181 tcttatacatttccagagattgacttatttgatgagatgcacttaccag 240
 QY 241 aactatactagagagatgacatgtgtggaactcattagactcagattccatcgt 300
 DB 241 aactatactagagagatgacatgtgtggaactcattagactcagattccatcgt 300
 QY 301 agcatttagagagatgacatgtgtggaactcattagactcagattccatcgt 360
 DB 301 agcatttagagagatgacatgtgtggaactcattagactcagattccatcgt 360

QY 361 aaacttgagaggttgctccttaaaaaattagatgcatctatacaacatccgtattaa 420
 DB 361 aaacttgagaggttgctccttaaaaaattagatgcatctatacaacatccgtattaa 420
 QY 421 aaatatattcattaccattaccatgaagtgctgtttgcagataatgagaagatgcatt 480
 DB 421 aaatatattcattaccattaccatgaagtgctgtttgcagataatgagaagatgcatt 480
 QY 481 gcagaatttgatcaataa 500
 DB 481 gcagaatttgatcaataa 500

RESULT 2

AAH20174
 ID AAH20174 standard; DNA: 12793 BP.
 XX
 AC AAH20174;
 XX
 DT 09-AUG-2001 (first entry)

DE Human spastin nucleotide sequence SEQ ID NO:1.

XX Human: mouse; spastin; ARSA; chromosome 13q11; identification;
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
 OS Homo sapiens.

FT Key Location/Qualifiers
 FT CDS 77..11566
 FT /tag= a
 FT /product= "spastin"

PN MO200129266-A2.

PD 26-APR-2001.

XX 20-OCT-2000; 2000MO-US29130.

XX 20-OCT-1999; 99US-0160588.

PA (UYMC-) UNIV MCGILL.
 PA (HOP1-) HOPITAL SAINTE-JUSTINE.

PI Hudson TJ, Engert J, Richter A;

DR WPI: 2001-308494/32.

XX P-PSDB: AAB97819.

PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -

PS Claim 1: Fig 9; 76pp: English.

CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSA) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce

CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (I). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with Spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes human spastin as given in the present invention.

XX
SQ Sequence 12793 BP; 4163 A; 2255 C; 2487 G; 3887 T; 0 other;

Query Match 100.0%; Score 500; DB 22; Length 12793;
Best Local Similarity 100.0%; Pred. No. 8.6e-127;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgattacaggaagaccatgactcagctcagcttcaatccagaacgatttcagc 60
DB 1 atgattacaggaagaccatgactcagctcagcttcaatccagaacgatttcagc 60
QY 61 tcttatcaaggaagtaatacatcttgcctgcgcagagaattgatttcaatgta 120
DB 61 tcttatcaaggaagtaatacatcttgcctgcgcagagaattgatttcaatgta 120
QY 121 tccattgatgaagaacagaatccaccatctgtttcatgcttaagatggttgaaaaa 180
DB 121 tccattgatgaagaacagaatccaccatctgtttcatgcttaagatggttgaaaaa 180
QY 181 tcttatacatcttttcagagagatttgactttttgataagatgacattatcccaag 240
DB 181 tcttatacatcttttcagagagatttgactttttgataagatgacattatcccaag 240
QY 241 aactatactagaagagtcagacatgtgtgaactcattagactcagattccatcgt 300
DB 241 aactatactagaagagtcagacatgtgtgaactcattagactcagattccatcgt 300
QY 301 agtcattttgagcagatgaatcagaacagcttcagaatttttagcagaatgtaca 360
DB 301 agtcattttgagcagatgaatcagaacagcttcagaatttttagcagaatgtaca 360
QY 361 aaaacttgagggttttccttaaaaattagatgcatctataaactccgcttataa 420
DB 361 aaaacttgagggttttccttaaaaattagatgcatctataaactccgcttataa 420
QY 421 aaaatattcatcaccattaccagaatgctgttttgagataatgagaagatgcatt 480
DB 421 aaaatattcatcaccattaccagaatgctgttttgagataatgagaagatgcatt 480
QY 481 gcagaaatgtgtatacaaa 500
DB 481 gcagaaatgtgtatacaaa 500

RESULT 3
AAH20178
ID AAH20178 standard; DNA: 12793 BP.

XX AAH20178;

XX 09-AUG-2001 (first entry)

DE Human mutated spastin nucleotide sequence SEQ ID NO:11.

XX Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
KM autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KM neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KM reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
KM atrophy of upper cerebellar vermis; absence of Purkinje cell;
KM abnormal neuronal lipid storage; genetic disorder; characterisation; ds.

OS Homo sapiens.
OS Synthetic.
XX WO200129266-A2.
XX 26-Apr-2001.
XX 20-Oct-2000; 2000WO-US29130.
XX 20-Oct-1999; 99US-0160588.
XX (UyMC-) UNIV MCGILL.
XX (HOPIT-) HOPITAL SAINTE-JUSTINE.
XX Hudson TJ, Engert J, Richter A;
XX WPI: 2001-308494/32.
XX
PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -
PS
PS Claim 1; Page -: 76pp; English.

CC The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
CC chromosome 13q11. (I) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (I) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (I). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with Spastin
CC protein or spastin gene expression or activity. The present sequence
CC represents a mutated human spastin gene from the present invention.
CC N.B. The present sequence is not given in the present specification but
CC is derived from the human spastin nucleotide sequence (AAH20174) as
CC stated on page 14.

XX
SQ Sequence 12793 BP; 4163 A; 2255 C; 2487 G; 3887 T; 0 other;

Query Match 100.0%; Score 500; DB 22; Length 12793;
Best Local Similarity 100.0%; Pred. No. 8.6e-127;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgattacaggaagaccatgactcagctcagcttcaatccagaacgatttcagc 60
DB 1 atgattacaggaagaccatgactcagctcagcttcaatccagaacgatttcagc 60
QY 61 tcttatcaaggaagtaatacatcttgcctgcgcagagaattgatttcaatgta 120
DB 61 tcttatcaaggaagtaatacatcttgcctgcgcagagaattgatttcaatgta 120
QY 121 tccattgatgaagaacagaatccaccatctgtttcatgcttaagatggttgaaaaa 180
DB 121 tccattgatgaagaacagaatccaccatctgtttcatgcttaagatggttgaaaaa 180
QY 181 tcttatacatcttttcagagagatttgactttttgataagatgacattatcccaag 240
DB 181 tcttatacatcttttcagagagatttgactttttgataagatgacattatcccaag 240

KM Human: mouse; spastin; ARSACS; chromosome 13q11; identification;
 KM autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KM neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KM reduced motor nerve velocity; hypermyelination of retinal nerve fibres;
 KM atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KM abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
 XX
 OS Homo sapiens.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT CDS 77..11566
 FT /*tag= a
 FT /product= "mutated spastin"
 FT
 PN W0200129266-A2.
 XX
 PD 26-APR-2001.
 XX
 PF 20-OCT-2000; 2000MO-US29130.
 XX
 PR 20-OCT-1999; 99US-0160588.
 XX
 PA (UYMC-) UNIV MCGILL.
 PA (HOP-) HOPITAL SAINTE-JUSTINE.
 XX
 PI Hudson TJ, Richter A;
 XX
 DR WPI: 2001-308494/32.
 DR P-PSDB: AAB97823.
 XX
 PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -
 PS
 PS Claim 1: Page -: 76pp; English.
 XX
 CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (II), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (II) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (II) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with Spastin
 CC protein or spastin gene expression or activity. The present sequence
 CC encodes a mutated human spastin.
 CC N.B. The present sequence is not given in the present specification but
 CC is derived from the human spastin nucleotide sequence (AAH20174) as
 CC stated on page 14.
 XX
 SO Sequence 12793 BP: 4163 A; 2257 C; 2487 G; 3886 T; 0 other;

Query Match 100.0%; Score 500; DB 22; Length 12793;
 Best Local Similarity 100.0%; Pred. No. 8.6e-127;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgattacagaagacatctactcagctcagcttctaataccagaacattgcacg 60

Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 atgattacagaagacatctactcagctcagcttctaataccagaacattgcacg 60
 Qy tcttatacagaagaatgaatacatctcgtcgtcgagagaattgattcaatgta 120
 Db tcttatacagaagaatgaatacatctcgtcgtcgagagaattgattcaatgta 120
 Qy tccatttgatgaagaacagaaatcccatctgttcatctggttaagaatgttgaa 180
 Db tccatttgatgaagaacagaaatcccatctgttcatctggttaagaatgttgaa 180
 Qy tcttatacatttctcagagagatttgacttatttgatagatgcccattcccccag 240
 Db tcttatacatttctcagagagatttgacttatttgatagatgcccattcccccag 240
 Qy 241 aactatacagaagaagtcagacatgtgtgaaactcattagactcaagatccatg 300
 Db 241 aactatacagaagaagtcagacatgtgtgaaactcattagactcaagatccatg 300
 Qy 301 agtcatttgacagatgaatctgagacagcttccagaatttttagagacattgta 360
 Db 301 agtcatttgacagatgaatctgagacagcttccagaatttttagagacattgta 360
 Qy 361 aaaacttgagaggttctccttaaaaaattagatgacatctacacatccgatttaa 420
 Db 361 aaaacttgagaggttctccttaaaaaattagatgacatctacacatccgatttaa 420
 Qy 421 aaatatattcattccaccattaccagatgctgttttcagagataatgagaatgcat 480
 Db 421 aaatatattcattccaccattaccagatgctgttttcagagataatgagaatgcat 480
 Qy 481 gcagaaatgtgtatcaaa 500
 Db 481 gcagaaatgtgtatcaaa 500
 Db 481 gcagaaatgtgtatcaaa 500

RESULT 6

AAH20175
ID AAH20175 standard; DNA: 11493 BP.

AC AAH20175;

DT 09-AUG-2001 (first entry)

XX Mouse spastin nucleotide sequence SEQ ID NO:3.

KM Human: mouse; spastin; ARSACS; chromosome 13q11; identification;
 KM autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KM neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KM reduced motor nerve velocity; hypermyelination of retinal nerve fibres;
 KM atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KM abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
 XX
 OS Mus musculus.

FH Key Location/Qualifiers
 FT CDS 1..11493
 FT /*tag= a
 FT /product= "spastin"
 PN W0200129266-A2.
 XX
 PD 26-APR-2001.
 XX
 PF 20-OCT-2000; 2000MO-US29130.
 XX
 PR 20-OCT-1999; 99US-0160588.
 XX

PA (UYMC-) UNIV MCGILL.
 PA (HOP-) HOPITAL SAINTE-JUSTINE.
 PI Hudson TJ, Richter A;

DR WPI: 2001-308494/32.
 DR P-PSDB: AAB97820.
 XX
 PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -
 XX
 PS
 XX Claim 1: Fig 8; 76pp; English.

CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARACS)) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with spastin
 CC protein or spastin gene expression or activity. The present sequence
 CC encodes mouse spastin as given in the present invention.
 XX

Sequence 11493 BP; 3599 A; 2281 C; 2387 G; 3226 T; 0 other;

Query Match 66.2%; Score 331.2; DB 22; Length 11493;
 Best Local Similarity 86.3%; Pred. No. 1.3e-80;
 Matches 366; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 77 atgaatacattcgtccgtcagagaatgtatgtcaatgtatccattgtatgaaac 136
 DB 1 atgaatacattcgtccgtcagagaatgtatgtcaggtatgtatccattgtatgaag 60
 QY 137 agaaatcaccatcgttcatgtcgttaagatgtttggaataattttatatacattt 196
 DB 61 aacggtcaccatcccttcaatggtcttaagaatgtttggaataattttatatacattt 120
 QY 197 tcagagatttgacttatttgatgagatgaccattatccccaagaactatactagagaa 256
 DB 121 tcggaagatttgacttatttgatgagatgaccattatccccaagaactatactagag 180
 QY 257 ggtcagacatgtgtggaactcaatgaactcagatccatcglttagtcatcttagagat 316
 DB 181 gaccagacgtgtgtggaactcagatcagatccatcagtagtcatcttagttagt 240
 QY 317 gaatcgaagcagagcttccagaatttttagcagacatgttacaanaaacttgaaggtt 376
 DB 241 gaacatgaagcagcttccagaatttttagcagacatgttacaanaaacttgaaggtt 300
 QY 377 gtccctaaanaatgaatcatcatcaacaacgcgtttatcaanaaataatactatcca 436
 DB 301 gtccctaaanaatgaatcatcatcaacaacgcgtttatcaanaaataatactatcca 360
 QY 437 ccattacccaagtgtgttttgagataatgagagaagatgcacattgagagaatgtgtaat 496
 DB 361 ccactcccgagtgctattttgacataatgagagaagataactctacagagaagttgttaat 420
 QY 497 caaa 500
 DB 421 caaa 424

RESULT 7
 ID AAA70119/C
 XX AAA70119 standard; DNA; 1251 BP.
 AC AAA70119;
 XX
 DT 07-NOV-2000 (first entry)
 XX
 DE Plasmodium falciparum chromosome 2 related DNA sequence SFO ID NO:252.
 XX
 XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
 KW antimalarial; malaria; protozoacide; infection; insecticide; ds.
 XX
 OS Plasmodium falciparum.
 XX
 PN WO200025728-A2.
 XX
 PD 11-MAY-2000.
 XX
 PF 05-NOV-1999; 99WO-US26796.
 XX
 PR 05-NOV-1998; 98US-0107131.
 XX
 PA (HOFF/) HOFFMAN S.
 PA (CARU/) CARUCCI D.
 PA (GARD/) GARDNER M.
 PA (VENT/) VENTER J C.
 XX
 PI Hoffman S, Carucci D, Gardner M, Venter JC;
 XX
 DR WPI: 2000-365347/31.
 XX
 PT Proteins encoded by chromosome 2 of the human malarial parasite,
 PT Plasmodium falciparum, useful as antimalarial vaccines and in the
 PT diagnosis of P.falciparum infection -
 XX
 PS Disclosure; Page 471; 577pp; English.

CC The present invention describes proteins and their fragments (I) encoded
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
 CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
 CC vaccines against P. falciparum infection comprising (I) or (II).
 CC (I) and (II) are useful for the development of vaccines against
 CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
 CC antibody raised to immunogens comprising the sequences of (I), are
 CC useful in the detection of infection with P. falciparum. Furthermore,
 CC (I) (especially when they are rifins or secreted or membrane proteins)
 CC can aid the identification of drugs to treat or prevent P. falciparum
 CC infection, or they can be used to identify drug resistance in
 CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
 CC subsequent identification of proteins encoded by it will help to expand
 CC our understanding of parasite biology, a process hampered by the
 CC complexity of the parasitic life cycle, and provide new targets for
 CC vaccine and drug development. Parasite resistance to drugs and mosquito
 CC resistance to insecticides have led to a resurgence of malaria in many
 CC parts of the world, and there is a pressing need for vaccines and new
 CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
 CC and protein sequences given in the present invention, but which are not
 CC specifically mentioned within the specification.
 XX

Sequence 1251 BP; 581 A; 93 C; 171 G; 406 T; 0 other;

Query Match 8.4%; Score 42.2; DB 21; Length 1251;
 Best Local Similarity 49.3%; Pred. No. 0.084;
 Matches 110; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 33 agcttctaataccagaagattgtcacgcttcttcaagaagcaatgaatacatcttcgc 92
 DB 690 AGGTTTAAATATATATTCATTCGATATGCAATTAATTCCTTCAATATAGCCCT 631
 QY 93 ctgcagagatgattgttcaatggtatcatcttgatgaaacagaataccaccatctg 152

Db 630 CTTTAAATTAATTAATCTACTGATCATTTTATGAAATAATATTATGATTTT 571
Qy 153 ttcaatgagcttaagaatggtttggaataaactttatatacatlctcagagagattgactt 212
Db 570 ATTATATAAATAAATATTTTATTAATATCATATCACTAATTAATGATTAATTTGTATA 511
Qy 213 tattgatgagatgcccactatccccagaactatactacagaga 255
Db 510 ATTATATATTTATTTTATTAATATAATTAATATGCTACTGTTGTA 468

RESULT 8
AAH32525/c
ID AAH32525 standard; cDNA; 1632 BP.
XX
AC AAH32525;
XX
DT 10-AUG-2001 (first entry)
XX
DE Human secreted protein-encoding gene 4 cDNA clone HCMH98, SEQ ID NO:14.
XX
KW Human; secreted protein; proliferative disorder; cancer; chromosome 11;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angioinetic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder; tumour;
KW endocrine disorder; infection; wound healing; vulnerrary;
KW cell culture; chemotaxis; food additive;
KW binding partner identification; ss.
XX
KW Homo sapiens.
OS
PN WO200134628-A1.
XX
PD 17-MAY-2001.
XX
PF 08-NOV-2000; 2000WO-US30653.
XX
PR 12-NOV-1999; 99US-0164735.
PR 27-JUL-2000; 2000US-0221193.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis GA, Birse CE, Ni J, Moore PA;
XX
DR WPI: 2001-339066/34.
DR P-PSDB; AAG73349.
XX
PT Nucleic acids encoding 35 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT disease and diabetic retinopathy -
XX
PS Claim 4; Page 444; 604pp; English.
XX
XX AAH32522-AAH32627 represent cDNAs corresponding to 35 human secreted
CC protein genes, and AAG73346-AAG73448 represent the proteins they encode.
CC AAG73449-AAG73519 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 52 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,

CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angioinetic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein-encoding cDNA of the invention.
XX
SQ Sequence 1632 BP; 375 A; 340 C; 387 G; 530 T; 0 other;

Query Match 7.7%; Score 38.4; DB 22; Length 1632;
Best Local Similarity 50.5%; Pred. No. 1;
Matches 93; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

Qy 26 cagctgcagctcttaaacccagaagattgcacgctctatcaagaaagtaatgataca 85
Db 1472 CAGATACATGATTTTAATCTAGACATTAACCCAGTTTAAATATTCTGGAANAC 1413
Qy 86 ttctgacctgcagagaatgtatgttcaatggtatccattgataagaaacagaatcaac 145
Db 1412 TATAGACTTTTACCTTATATGACCTTTCAAAGGTACTATGATTTGTAATAATTCAGTTT 1353
Qy 146 ccaatcgtttatgcttaagaatgttgcgaaatactttatatacatlctcagagat 205
Db 1352 CCACTGTGTACATGATATATCTAGTCCATTAATCCAGTAATCTGTTTCATGATGAC 1293
Qy 206 ttga 209
Db 1292 CTCA 1289

RESULT 9
AAS02544/c
ID AAS02544 standard; cDNA; 1892 BP.
XX
AC AAS02544;
XX
DT 18-JUL-2001 (first entry)
XX
DE Human secreted protein gene #25.
XX
KW Human secreted protein; diagnosis; autoimmune disease; ss;
KW rheumatoid arthritis; hyperproliferative disorder; neoplasm; sunburn;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder; skin aging;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; chemotaxis; preservative;
KW organ transplantation; tissue regeneration; food additive.
XX
KW Homo sapiens.
OS
PN WO200123409-A2.
XX
PD 05-APR-2001.
XX
PF 26-SEP-2000; 2000WO-US26371.
XX
PR 27-SEP-1999; 99US-0155804.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis GA;
XX
DR WPI: 2001-266139/27.
DR P-PSDB; AAU01636.

XX Nucleic acids encoding 38 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT disease and diabetic retinopathy -
XX
XX
PS Disclosure: Page 435; 488pp; English.
XX
XX AAS02511-AAS02557 represent human secreted protein coding sequences
CC and primers of the invention. The human secreted protein sequences are
CC used to prevent, treat or ameliorate a medical condition in e.g. humans,
CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are
CC also used in diagnosing a pathological condition or susceptibility to a
CC pathological condition. The antibodies to human secreted proteins can
CC also be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. Rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. The polypeptides can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities.
XX
XX Sequence 1892 BP; 447 A; 371 C; 478 G; 593 T; 3 other;

Query Match	7.7%	Score 38.4	DB 22,	Length 1892,
Best Local Similarity	50.5%	Pred: NO. 1.1,		
Matches 93, Conservative	0,	Mismatches 91,	Indels 0,	Gaps 0

QY	26	cagctgagcgttccaattccagaagattgcacgctttacaaggaaagttaatgatca	85
Db	1731	CAGTACATCATTTAAATCTAGCACATAAACCAGTTTAAGAATTATTTTGTGAAAAC	1672
QY	86	tctcgccctggcagaagaattgattgttcactgatacatttgatgaacaacagaatcac	145
Db	1671	TATAGACTTTTACCCCTTAGACCTTTCAAGSGPACCTTACATTGTATAATTTCAGTTTT	1612
QY	146	ccactgcttcatcgctctaagaatglttgcgaaaaccttatatacatctttcagaggat	205
Db	1611	CCACTGTTCAATAGATATATACAGTGCCATTAATCCATAATCTGTTTCATGATGAC	1552
QY	206	ttga	209
Db	1551	CTCA	1548

RESULT	ID
10	ABL12784/c
	ABL12784 standard; cDNA; 3644 BP.

DT 26-MAR-2002 (first entry)

DE *Drosophila melanogaster* expressed polynucleotide SEQ ID NO 32834.

KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.

Drosophila melanogaster.

PN WO200171042-A2

PD 27-SEP-2001

PF 23-MAR-2001; 2001WO-US09231.

XX

PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
...

DR WPI; 2001-656860/75.
DR P-PSDB; ABB68681.

PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from *Drosophila* and for elucidating cell signalling and cell-cell

PS Claim 1; SEQ ID NO 32834; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB161716-AB130511), expressed DNA
CC sequences (AB101840-AB161675) and, the encoded proteins
CC (AAB57737-AB272072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.

SQ Sequence 3644 BP; 1243 A; 728 C; 721 G; 952 T; 0 other;

Query Match	7.7%	Score 38.4;	DB 23;	Length 3644;
Best Local Similarity	47.8%	Pred. No. 1.3;		
Matches 111; Conservative	0;	Mismatches 121;	Indels 0;	Gaps 0

QY	215	tttaatgagatgccactatcccaagacatacttaggaagaagtgaacatgtgtgaa	274
Db	362	tcctgacacgatttttctgtgctcttcagatcttggaaatgacgacgtgcgaatttcgaagcc	303
QY	275	ctcatagactcagatctcactcgcttagcatctttagacgatgaatctgaagcagctt	334
Db	302	ctttatgtgcgcttgaatgatgccccaagaacatgctctgcactatgaagccgcttcag	243
QY	335	ccgaacattttagcagacatgtgtcaaaaacttggagggtctgtcccttaaaaaatgat	394
Db	242	ccagatttggtaactggaattggaaaagaagtgggcttgaagattttttatataaaatctgg	183
QY	395	gcactctacaacatccgcttatataaanaatatatcatccaccattacca	446
Db	182	gaatgagattttacataaaaatttttaatttaattgaattgttttaaacaa	131

RESULT	11
AAI61373/c	
ID	AAI61373 standard; DNA; 513445 BP.

DT 16-OCT-2001 (first entry)

DE Soybean 318013 region A3, SEQ ID NO: 4

KW Soybean; antihelminthic; gene therapy; soybean cyst nematode; SCN;
SCN resistance; rhg1; Rhg4; SCN resistant allele; plant breeding;;
KW

XX

XX

XX

XX

PE

XX 07-JAN-2000; 2000US-0174880.
XX (MONS) MONSANTO CO.
XX Hauge BM, Wang ML, Parsons JD, Parnell LD;
XX WPI: 2001-425872/45.
XX P-PSDB; AAM42216.
XX New purified nucleic acid for producing a soybean plant having soybean
XX cyst nematode resistance and for use in plant breeding programs -
XX
XX Claim 30; Page 596-893; 1353pp; English.
XX The invention relates to nucleic acid molecules from regions of the
XX soybean genome which are associated with soybean cyst nematode (SCN)
XX resistance. The nucleic acids are used to transform plants, and can
XX produce soybean plants having an rhg1 or an Rhg4 SCN resistant allele.
XX The nucleic acids can be used for investigating rhg1 or Rhg4 haplotypes
XX of soybean plants and for introgressing SCN resistance or partial SCN
XX resistance into soybean plants. They can also be used in plant breeding
XX programmes. The invention also relates to proteins encoded by such
XX nucleic acid molecules, as well as antibodies capable of recognising
XX these proteins. The present sequence is a nucleic acid molecule
XX provided in the specification.
XX
SQ Sequence 513445 BP; 173367 A; 85402 C; 83912 G; 170492 T; 272 other;

Query Match 7.6%; Score 37.8; DB 22; Length 513445;
Best Local Similarity 48.4%; Pred. No. 9.8; Mismatches 112; Indels 0; Gaps 0;
Matches 105; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 178 aaacttataatactttcagagagattgacttatttgatgagatgcacattccc 237
DB 477367 AATTATGTAATAATATATAGATTAAAGAAAAATAAATGATGCTAATGATGATAA 477308
QY 238 cagagacatactagagagagatgcacatgctgctggaactcattgagatcagatccatc 297
DB 477307 ATATATATTTATATACATATATATATTTATATATAGTTATATATATATATATTCATC 477248
QY 298 gttagcaatttagacatgacatgcagacacagcttcagagaattttagacagacatgct 357
DB 477247 ATTAATCATATATATATGATCAATTTACTGAACTTATATAAAAAACTTAAAAAAGT 477188
QY 358 acaaaacttgagaggttgcctctaaataatgat 394
DB 477187 ACATTATGTTGATTTTATTTATGATCGACATTGTAAT 477151

RESULT 12
ABL26428/c
ID ABL26428 standard; DNA: 7957 BP.
XX
XX ABL26428;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 30757.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
XX
XX OS
XX PN WO200171042-A2.
XX
XX PD 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX

PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI: 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1; SEQ ID NO 30757; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (AAB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 7957 BP; 2081 A; 1836 C; 1817 G; 2223 T; 0 other;

Query Match 7.5%; Score 37.4; DB 23; Length 7957;
Best Local Similarity 53.0%; Pred. No. 3.2; Mismatches 71; Indels 0; Gaps 0;
Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 99 gagaatgatgttcaatggtatccattgagaaacagaatacaccatgttcat 158
DB 7286 GATAATCAATTTTGTGGAATCTATTTAATCAGATTAAATGAAATGAAATGAAAT 7227
QY 159 ggttaagatggttggaanaatccttatatacatctttcagagatctgacttattg 218
DB 7226 TGATTAGTTACATATGTAATAATATTTACAGCAATTTTATTAATTAATGTTACTTCG 7167
QY 219 atgagatgcaccttaccacgaactatc 249
DB 7166 AFAATTACCACTTCACAAAGTTATATAT 7136

RESULT 13
ABL34053/c
ID ABL34053 standard; DNA: 11172 BP.
XX
XX ABL34053;
XX
XX 26-MAR-2002 (first entry)
XX
XX Human immune system associated gene SEQ ID NO: 2026.
XX
XX Human; immune system disease; cytosine methylation; antiasthmatic;
XX antiarteriosclerotic; antianaemic; cytosolic; nootropic;
XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX antirheumatic; antiarthritic; antidiabetic; antiporiatic;
XX antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
XX gene; ds.
XX
XX Homo sapiens.
XX
XX OS
XX PN WO200200928-A2.
XX
XX PD 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-EP07537.
XX

xx 18-JUN-1999; 99US-0140121.
xx
xx (INCY-) INCYTE GENOMICS INC.
xx
xx Lagace RE, Patterson C, Berg KL;
xx WPI. 2001-041427/05.
xx
xx Genomic library for identifying diagnostic and therapeutic
xx compositions, and for identifying virulence factors, regulatory
xx elements and drug targets, comprises Moraxella catarrhalis nucleic
xx acids -
xx
xx Claim 1; Page 171-180; 545pp; English.
xx
xx The present invention relates to a Moraxella catarrhalis genomic library
xx comprising of a combination of 41 nucleic acid molecules (see
xx AAF28514-AAF28554). The library has a number of uses described in the
xx specification e.g. is useful for identifying diagnostic and therapeutic
xx compositions. M. catarrhalis (Branhamella catarrhalis) is a large
xx aerobic, gram-negative diplococcus, normally found among the bacterial
xx flora of human upper airways. M. catarrhalis is known to cause acute,
xx localised infections such as otitis media, sinusitis and bronchopulmonary
xx infection and life-threatening, systemic diseases including endocarditis
xx and meningitis.
xx
xx Sequence 39003 BP; 11568 A; 8751 C; 7476 G; 11208 T; 0 other;

Query Match	7.4%	Score 37.2	DB 22	Length 39003
Best Local Similarity	52.6%	Pred. NO. 6.1		
Matches	81	Conservative	0	Mismatches 73; Indels 0; Gaps 0;
Qy	291	ttccatcgttga	tcattttaagcagatcgaacgaacgtccagaattttgacag	350
Db	22598	ttatctgcttgtagc	atcaaacactgtgttttaagaacactgtcccaagtttaagaag	22657
Qy	351	acattgtacaaa	aactcttggaaggtttgtccttaaaaaaatttagatcctatcacacatc	410
Db	22658	aagatgtttaa	aattcctataaaaaatccttaataaaaaatccttaataaaaaaata	22717
Qy	411	cgcctatcaaaa	atatatttcacatcacc	444
Db	22718	tccttaataaaa	aatatccttaataaaaaatc	22751

Search completed: May 22, 2002, 06:43:32
Job time: 8022 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 04:28:15 ; Search time 108.82 Seconds
(without alignments)
1128.623 Million cell updates/sec

Title: US-09-693-205-7_COPY_1_500
Perfect score: 500
Sequence: 1 atgattcagaagaagacat.....gcagaattgttaatacaaa 500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/p10data/1/ina/5A_COMB.seq:*
2: /cgn2_6/p10data/1/ina/5B_COMB.seq:*
3: /cgn2_6/p10data/1/ina/6A_COMB.seq:*
4: /cgn2_6/p10data/1/ina/6B_COMB.seq:*
5: /cgn2_6/p10data/1/ina/PCRTUS_COMB.seq:*
6: /cgn2_6/p10data/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	35.6	7.1	3129	4 US-09-387-695-1	Sequence 1, Appl
C 2	34.8	7.0	658	1 US-09-328-111-775	Sequence 775, App
C 3	34.2	6.8	3456	1 US-08-190-687B-24	Sequence 24, Appl
C 4	34.2	6.8	4307	1 US-08-190-687B-7	Sequence 7, Appl
C 5	33.2	6.6	6479	2 US-08-649-046-3	Sequence 3, Appl
C 6	32.8	6.6	853	2 US-08-179-557-19	Sequence 19, Appl
C 7	32.4	6.5	3000	1 US-08-680-395-1	Sequence 1, Appl
C 8	32	6.4	149	1 US-08-629-339-4	Sequence 4, Appl
C 9	32	6.4	149	1 US-08-759-873-4	Sequence 4, Appl
C 10	32	6.4	180	3 US-09-004-113-13	Sequence 13, Appl
C 11	32	6.4	11832	2 US-08-416-603-1	Sequence 1, Appl
C 12	31.8	6.4	1419	1 US-08-242-098-39	Sequence 39, Appl
C 13	31.6	6.3	2016	1 US-09-132-118-1	Sequence 1, Appl
C 14	31.6	6.3	2106	4 US-08-923-511-1	Sequence 1, Appl
C 15	31.6	6.3	2137	1 US-08-444-005-16	Sequence 16, Appl
C 16	31.6	6.3	2617	3 US-09-161-443-1	Sequence 1, Appl
C 17	31.6	6.3	9687	4 US-09-133-944-2	Sequence 2, Appl
C 18	31.4	6.3	5014	4 US-09-381-862-3	Sequence 3, Appl
C 19	31	6.2	2848	4 US-08-936-165A-197	Sequence 197, App
C 20	31	6.2	2897	2 US-08-927-394-1	Sequence 1, Appl
C 21	31	6.2	3923	4 US-09-176-320-7	Sequence 7, Appl
C 22	31	6.2	6559	4 US-09-234-186-1	Sequence 1, Appl
C 23	31	6.2	6560	5 PCT-US93-05651-1	Sequence 1, Appl
C 24	30.8	6.2	854	3 US-08-867-381A-51	Sequence 51, Appl
C 25	30.8	6.2	854	4 US-09-521-144-51	Sequence 43, Appl
C 26	30.8	6.2	1002	4 US-08-960-780-43	Sequence 43, Appl
C 27	30.8	6.2	1002	4 US-09-073-898-43	Sequence 43, Appl

28	30.8	6.2	1500	1 US-08-117-083-67	Sequence 67, Appl
29	30.8	6.2	2239	4 US-09-196-390-1	Sequence 1, Appl
30	30.8	6.2	3000	1 US-08-184-252A-1	Sequence 1, Appl
31	30.8	6.2	3000	5 PCT-US95-00601-1	Sequence 1, Appl
32	30.6	6.1	9636	1 US-08-323-170B-1	Sequence 1, Appl
33	30.6	6.1	9636	4 US-08-954-441-1	Sequence 1, Appl
34	30.4	6.1	1491	2 US-08-218-265-9	Sequence 9, Appl
35	30.2	6.0	2750	3 US-08-617-860B-33	Sequence 33, Appl
36	30	6.0	4619	2 US-08-874-186-38	Sequence 38, Appl
37	29.8	6.0	13158	2 US-08-687-080-105	Sequence 105, App
C 38	29.6	5.9	4447	2 US-08-304-309-3	Sequence 3, Appl
C 39	29.6	5.9	40352	3 US-08-991-942-3	Sequence 15, Appl
C 40	29.6	5.9	4447	3 US-08-846-111D-15	Sequence 1, Appl
C 41	29.6	5.9	56516	2 US-08-996-306-1	Sequence 1, Appl
C 42	29.6	5.9	56516	4 US-09-338-907-1	Sequence 1, Appl
C 43	29.6	5.9	56516	4 US-09-218-207-1	Sequence 1, Appl
C 44	29.6	5.9	56520	4 US-09-338-907-179	Sequence 179, App
C 45	29.6	5.9	56520	4 US-09-218-207-179	Sequence 179, App

ALIGNMENTS

```
RESULT 1
US-09-387-695-1/c
: Sequence 1, Application US/09387695
: Patent No. 6280990
: GENERAL INFORMATION:
: APPLICANT: May, Earl
: APPLICANT: Van Horn, Stephanie
: APPLICANT: Warren, Patrick V.
: APPLICANT: Warren, Richard L.
: TITLE OF INVENTION: dnae
: FILE REFERENCE: GM10237
: CURRENT APPLICATION NUMBER: US/09/387, 695
: CURRENT FILING DATE: 1999-08-31
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1
: LENGTH: 3129
: TYPE: DNA
: ORGANISM: Streptococcus pneumoniae
: US-09-387-695-1

Query Match 7.1%; Score 35.6; DB 4; Length 3129;
Best Local Similarity 51.9%; Pred. No. 0.5;
Matches 80; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 106 gatgttcataatgatatcatcattgattgaaacagaaatccaccatctgtttatgcttae 165
DB 2543 GATTCCTGCGCCACATTAATAGCATCTCCAAACAACCTCCCAACTCTTTCACAAATTC 2484

QY 166 gatgttggaataatcttatcatcttttcagagagattgacttatgtatgagat 225
DB 2483 AATGATTAGCTAATTAATTAATCTTTTGACGAGATTCTTTTCAATGATCAGAAAGA 2424

QY 226 gccactatccccaactactactagaggaagt 259
DB 2423 CCAACTTTTACCAAGGTTTCAGCAGAGAGAGT 2390

RESULT 2
US-09-328-111-775
: Sequence 775, Application US/09328111
: Patent No. 6262333
: GENERAL INFORMATION:
: APPLICANT: Endege, Wilson O.
: APPLICANT: Steinmann, Kathleen E.
: APPLICANT: Astle, Jon H.
: APPLICANT: Burgess, Christopher C.
: APPLICANT: Bushnell, Steven E.
: APPLICANT: Carroll III, Eddie
```

APPLICANT: Catino, Theodore J.
APPLICANT: Dertl, Adnan
APPLICANT: Ford, Donna M.
APPLICANT: Lewis, Marcia E.
APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCD-257 (US)
CURRENT APPLICATION NUMBER: US/09/328,111
CURRENT FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: US 60/088,801
EARLIER FILING DATE: 1998-06-10
NUMBER OF SEQ ID NOS: 850
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 775
LENGTH: 658
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)-(658)
OTHER INFORMATION: n = A,T,C or G
US-09-328-111-775

Query Match 7.0%; Score 34.8; DB 4; Length 658;
Best Local Similarity 47.0%; Pred. No. 0.48;
Matches 108; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

QY 193 ttttcagaggaattgacttatttgatgagatgccactatccagaaactatactaga 252
DB 139 tttttatccatattagtcctgtgtagcattcttcaactaaccgctgtttgaaga 198
QY 253 ggaaggtcagactggtgtaactcttagaccagagatccctcgttagcttttga 312
DB 199 tgaatgatataccgacggttgaaggaagtgtagatccctatcgagataagaataaca 258
QY 313 cgaatgatccgaagcagcttcacgaatttttagcagacattgtacaacttggaag 372
DB 259 tgaagagctcagacttaagatgaggaatgatgtggccatataatgaactaga 318
QY 373 gtttgcttcaaaaatagatgcatctacacacatccgcttataaaa 422
DB 319 ggtggtcgtgagagagagagataaaaactcaaacctggaatatgatatataa 368

RESULT 3
US-08-190-687B-24
Sequence 24, Application US/08190687B
Patent No. 5760203

GENERAL INFORMATION:

APPLICANT: Wong, Gail L.
APPLICANT: Martin, George
APPLICANT: McCormick, Francis P.
APPLICANT: Rubinfeld, Bonnie
APPLICANT: O'Rourke, Edward C.
APPLICANT: Clark, Robin
TITLE OF INVENTION: GAP Gene Sequences
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/190,687B
FILING DATE: 02-FEB-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/774,644
FILING DATE: 11-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/260,807
FILING DATE: 21-OCT-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/230,761
FILING DATE: 10-AUG-1988
ATTORNEY/AGENT INFORMATION:
NAME: Gass, David A.
REGISTRATION NUMBER: 38,153
REFERENCE/DOCKET NUMBER: 27527/31898
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 3456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 100..2709
US-08-190-687B-24

Query Match 6.8%; Score 34.2; DB 1; Length 3456;
Best Local Similarity 55.5%; Pred. No. 1.3;
Matches 66; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 343 ttttagcagacattgtacaacttgaggggttgccttaaaaaatagatgcatctat 402
DB 2100 ttttaacacacacttttgacacttcttcagagccttgaggagaaatatttcacgcttcacaga 2159
QY 403 acacacatcgccttataaaaatataatcattaccattaccagaatgctgttttcaga 461
DB 2160 aatgacttccacgacacattgagatattatttggttttacagaaatcgttcacacata 2218

RESULT 4
US-08-190-687B-7
Sequence 7, Application US/08190687B
Patent No. 5760203

GENERAL INFORMATION:

APPLICANT: Wong, Gail L.
APPLICANT: Martin, George
APPLICANT: McCormick, Francis P.
APPLICANT: Rubinfeld, Bonnie
APPLICANT: O'Rourke, Edward C.
APPLICANT: Clark, Robin
TITLE OF INVENTION: GAP Gene Sequences
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/190,687B

FILING DATE: 02-FEB-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/774,644
FILING DATE: 11-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/260,807
FILING DATE: 21-OCT-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/230,761
FILING DATE: 10-AUG-1988
ATTORNEY/AGENT INFORMATION:
NAME: Gass, David A.
REGISTRATION NUMBER: 38,153
REFERENCE/DOCKET NUMBER: 27527/31898
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ. ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4307 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 119..3259
US-08-190-687B-7

Query Match 6.6%; Score 34.2; DB 1; Length 4307;
Best Local Similarity 55.5%; Pred. No. 1.4;
Matches 66; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
Qy 343 tttagcagacatgttacaacacttgagggttgctccttaaaatagatgcatcat 402
Db 2650 ttttaacacacctattgaacacttgcagaccttgagagaaattttcagcttcaga 2709
Qy 403 acaacaccgcttataaaaatatattcaccattaccacaaagtgcgtttgcaga 461
Db 2710 aattcttccacccgacattgagattattttatgsgcttttaccagaaatctgttcagcaca 2768
RESULT 5
US-08-649-046-3/c
Sequence 3, Application US/08649046
Patent No. 5912415
GENERAL INFORMATION:
APPLICANT: OLSEWSKI, NETL E.
APPLICANT: JACOBSEN, STEVEN E.
TITLE OF INVENTION: THE SPINDLY GENE, METHODS OF
TITLE OF INVENTION: IDENTIFICATION AND USE
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: MUEHLING, RAASCH, GEBHARDT & SCHWAPPACH, P.A.
STREET: 119 NORTH FOURTH STREET, SUITE 203
CITY: MINNEAPOLIS
STATE: MINNESOTA
COUNTRY: USA
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/649,046
FILING DATE: 16-MAY-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MCCORMACK, MYRA H.

REGISTRATION NUMBER: 36,602
REFERENCE/DOCKET NUMBER: 110.00340101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1225
TELEFAX: 612-305-1228
INFORMATION FOR SEQ. ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 6479 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-649-046-3

Query Match 6.6%; Score 33.2; DB 2; Length 6479;
Best Local Similarity 55.1%; Pred. No. 3.2;
Matches 65; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
Qy 382 taaaaattagatgcatctatacaacatccgcttataaaatatatcttcacatt 441
Db 4476 ttaaaatttaagatttttttttgagaccttccgaaccgaatttcattgaaccgaa 4417
Qy 442 accaagtcgttttgacagataatgagaaagatgccattgcagaattgttaacaa 499
Db 4416 acgtactttttgttgatctgttgatcagcaccagcgaatctgtaattctgtaacaa 4359
RESULT 6
US-08-179-557-19/c
Sequence 19, Application US/08179557
Patent No. 5837509
GENERAL INFORMATION:
APPLICANT: ISRAELSEN, Hans
APPLICANT: BECH HANSEN, Egon
APPLICANT: MADSEN, Soeren Michael
APPLICANT: JOHANSEN, Eric
APPLICANT: NILSSON, Dan
APPLICANT: VRANG, Astrid
TITLE OF INVENTION: Recombinant Lactic Acid Bacterium
TITLE OF INVENTION: Containing an inserted Promoter and Method of Constructing
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/179,557
FILING DATE: 07-JAN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 1579/92
FILING DATE: 30-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 0988/93
FILING DATE: 01-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DK94/00004
FILING DATE: 03-JAN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/036,681
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768


```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/242,098
FILING DATE: 13-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/133,390
FILING DATE: 08-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30307/141/PLVI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136

INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 1419 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-242-098-39

Query Match 6.4%; Score 31.8; DB 1; Length 1419;
Best Local Similarity 52.7%; Pred. No. 4.6;
Matches 69; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 315 atgaatctgaagcacagctccagaatttttagcacagatgtaacaaaacttgaaggt 374
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 707 AAGATTGTGAGAGCTGCACCTTGAAAATTTGTAGCTGGAACATGCAAGCAAAAATCTTCCTT 766

QY 375 ttgtcttaaaaaattagatgcatctatcacacaatcgcgttatataaaatatatcatt 434
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 767 TTGACCTTGAAGAGGCTGATGGAAGCAACACATTACGCTTATGTGTAATACGCGCATG 826

QY 435 caccatlacca 445
      ||| |||
Db 827 GTGCGATGCCA 837

RESULT 13
US-09-132-118-1
: Sequence 1, Application US/09132118
: Patent No. 6211337
GENERAL INFORMATION:
APPLICANT: BAICHMAL, VIJAY R
APPLICANT: HUANG, JIANING
APPLICANT: HSU, HAILING
APPLICANT: GOEDEL, DAVID V
TITLE OF INVENTION: RIP: NOVEL HUMAN PROTEIN INVOLVED IN
TITLE OF INVENTION: TUMOR NECROSIS FACTOR SIGNAL TRANSDUCTION, AND SCREENING
TITLE OF INVENTION: ASSAYS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 75 DENISE DRIVE
CITY: HILLSBOROUGH
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94010

COMPUTER READABLE FORM:

```



```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/132,118
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A.
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: 795-006-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 343-4341
TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO.: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2016 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cdna
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2013
US-09-132-118-1

```

Query Match	6.3%	Score 31.6	DB 4	Length 2016
Best Local Similarity	53.2%	Pred. No. 5.9		
Matches 67	Conservative 0	Mismatches 59	Indels 0	Gaps 0

[illegible]

QY	464	atgga	469
Db	952	ATGCAG	957

RESULT 14
US-08-923-511-1
; Sequence 1, Application US/08923511
; Patent No. 6074376

GENERAL INFORMATION:

1
 2 APPLICANT: Burnham, Martin K.
 3 APPLICANT: Fosseberry, Andrew
 4 APPLICANT: Hodgson, John
 5 APPLICANT: Lawlor, Elizabeth
 6 APPLICANT: Rosenberg, Martin
 7 APPLICANT: Ward, Judith
 8
 9 TITLE OF INVENTION: NO. 6274376el c1pL
 10
 11 NUMBER OF SEQUENCES: 5
 12
 13 CORRESPONDENCE ADDRESS:
 14 ADDRESS: Dechert, Price & Rhoads
 15 STREET: 4000 Bell Atlantic Tower, 1717 Arch Street
 16 CITY: Philadelphia

```

1 ZIP: 19103-2793
2
3 COMPUTER READABLE FORM:
4
5 MEDIUM TYPE: Diskette
6
7 COMPUTER: IBM Compatible
8
9 OPERATING SYSTEM: DOS
10
11 SOFTWARE: FASTEO for Windows Version 2.0
12
13 CURRENT APPLICATION DATA:
14
15 APPLICATION NUMBER: US/08/923,511
16
17 FILING DATE:
18

```

1 CLASSIFICATION: 435
 2 PRIOR APPLICATION DATA:
 3 APPLICATION NUMBER: 60/011888
 4 FILING DATE: 20-FEB-1996
 5 APPLICATION NUMBER: PCT/US97/025474
 6 FILING DATE: 19-FEB-1997
 7 APPLICATION NUMBER: PCT/US97/02318
 8 FILING DATE: 19-FEB-1997
 9 ATTORNEY/AGENT INFORMATION:
 10 NAME: Dickinson, O. Todd
 11 REGISTRATION NUMBER: 28,354
 12 REFERENCE/DOCKET NUMBER:
 13 TELECOMMUNICATION INFORMATION:
 14 TELEPHONE: 215/994-2252
 15 TELEFAX: 215/994-2222
 16 TELEX:
 17 INFORMATION FOR SEQ ID NO: 1:
 18 SEQUENCE CHARACTERISTICS:
 19 LENGTH: 2106 base pairs
 20 TYPE: nucleic acid
 21 STRANDEDNESS: single
 22 TOPOLOGY: linear
 23
 24 US-08-923-511-1

Query Match	6.3%;	Score 31.6;	DB 4;	Length 2106;
Best Local Similarity	60.5%;	Pred. No. 6;		
Matches 52;	Conservative 0;	Mismatches 34;	Indels 0;	Gaps 0

Oy	283	actcaggatctacgcgttatgctcttttagacgatcatctgaagcacagcctccagaatt	342
Db	1557	ACGCGTAATCCACTACACTCATCTCATTATTGTGAATAAAGCAAAATCCACAAT	1616
Oy	343	tttagcagacatctgtacaaaacttg	368
Db	1617	TTTAACATTGTTATTACAACTAATGG	1642

Db 1617 TTTACATTGTTATTACAAGTAATCG 1642

RESULT 15
US-08-444-005-16
; Sequence 16, Application US/08444005

GENERAL INFORMATION

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street, Suite 3100
CITY: Boston

COUNTRY: USA

```

? ZIP: 02110-2804
? COMPUTER READABLE FORM:
? MEDIUM type: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent In Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/444,005

```

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,164
REFERENCE/DOCKET NUMBER: 00383/0260001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 04:23:20 ; Search time 3619.39 Seconds
(without alignments)
1864.535 Million cell updates/sec

Title: US-09-693-205-7_COPY_1_500
Perfect score: 500
Sequence: 1 atgattcaggaagacatc.....gcagaattgttaatacaaa 500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:*
1: em_estda:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_iny:*
15: em_gss_pln:*
16: em_gss_vtc:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	396.6	79.3	746	12	BH025017
C 2	396.6	79.3	746	12	BH126217
C 3	230.6	46.1	558	12	AZ649875
C 4	212.4	42.5	241	9	BE008891
C 5	163.2	32.6	496	12	AZ644393
C 6	159.2	31.8	717	12	BH110525
C 7	142	28.4	883	12	CNS04HOG
C 8	83.2	16.6	202	9	BB561400
C 9	66.4	13.3	170	10	BF397419
C 10	57	11.4	266	10	BF397419
C 11	44	8.8	752	12	BH571801
C 12	43.8	8.8	440	12	AZ066617
C 13	41.4	8.3	895	12	BH131081
C 14	41.4	8.3	896	12	BH136596
C 15	41.4	8.3	940	12	AZ547902
C 16	41.4	8.3	1101	12	CNS0178M
C 17	41	8.2	1101	12	CNS00DDH

C 18	39.8	8.0	845	12	AZ679734
C 19	39.8	8.0	926	12	BH146623
C 20	39.6	7.9	653	12	AZ611614
C 21	39.6	7.9	753	12	AZ660635
C 22	39.6	7.9	928	12	AZ528035
C 23	39.6	7.9	987	12	CNS014PQ
C 24	39.4	7.9	642	9	BB523933
C 25	39.2	7.8	1026	12	CNS0122K
C 26	39	7.8	649	12	AO576178
C 27	38.8	7.8	871	12	BH133334
C 28	38.8	7.8	1101	12	CNS00F4N
C 29	38.6	7.7	299	10	BE578331
C 30	38.6	7.7	451	10	BM027431
C 31	38.4	7.7	387	9	AI275161
C 32	38.4	7.7	420	12	AQ003378
C 33	38.4	7.7	560	10	BF510708
C 34	38.4	7.7	804	10	BF220200
C 35	38.2	7.6	766	12	BH495660
C 36	38.2	7.6	877	12	AZ671825
C 37	37.8	7.6	165	9	AT544786
C 38	37.8	7.6	541	10	BI743087
C 39	37.4	7.5	418	9	AU020427
C 40	37.4	7.5	683	9	AI292912
C 41	37.4	7.5	776	10	BF294246
C 42	37.4	7.5	883	12	BH256414
C 43	37.4	7.5	932	12	BH155613
C 44	37.4	7.5	1101	12	CNS0039G
C 45	37.2	7.4	346	10	BF709704

ALIGNMENTS

RESULT 1
BH025017/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

house mouse.
Mus musculus

REFERENCE
AUTHORS

TITLE
JOURNAL
COMMENT

Other-GSSS: RPCI-24-318E15.TVB

Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@tigr.org). Clones may be purchased from BACPAC Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC end page: http://www.tigr.org/tigr/bac_ends/mouse/bac_end_intro.html
Plate: 318 row: E column: 15
Seq primer: SP6
Class: BAC ends.

FEATURES
Source

Location/Qualifiers
1..746
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-318E15"

RESULT 3
 AZ649875
 LOCUS 558 bp DNA linear GSS 14-DEC-2000
 DEFINITION 1M0519B14R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC1M0519B14 R. DNA sequence.
 ACCESSION AZ649875
 VERSION AZ649875.1 GI:11783794
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 558)
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,F., Reilly,
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
 and Wright,D., Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 JOURNAL Contact: Robert B. Weiss
 University of Utah
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SIC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0519 row: B column: 14
 Seq primer: CACACAGAGAAACAGCATGAC
 Class: plasmid ends
 High quality sequence stop: 558.
 Location/Qualifiers
 1..558
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0519B14"
 /clone_1fb="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: pMD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pMD42 (g114732114|p|AR129072.1) a copy-number
 inducible derivative of plasmid RL. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."
 BASE COUNT 200 a 99 c 96 g 163 t
 ORIGIN
 Query Match 46.1%; Score 230.6; DB 12; Length 558;
 Best Local Similarity 84.1%; Pred. No. 3e-50;
 Matches 297; Conservative 0; Mismatches 49; Indels 7; Gaps 3;
 QY 149 tctgttcacgcttaagcgttttgagaaatcttatatacatcttcagagattg 208
 11 |||||||||||||||| |||||| |||||||||||| |||||
 Db 1 TCCCTTCATGCGCTTAAGA----TTGGAAGAACTCTATATACATTTCTCAGAGATTG 56

QY 209 acttatttgatgagatgcacacttatccccaagaactacagaggaagcagacatgt 268
 ||||||||||||||||||||| ||||| ||| ||| |||
 Db 57 ACTTTATTGATGAGATGCCACTTATCCCTTGAAATATATACGATGAGACCAGATGTGT 116
 QY 269 gtggaacttaagactcagaagatccatcgtctagtcattttagacgatactgaagca 328
 ||||||||| ||||||||| ||||| ||||||||| ||||| |||||
 Db 117 GTGGAACATCAGACAGCTCAGATCCATCAGTACGATTTTAAAGATGAAGCAAGTCAACT 176
 QY 329 cagcttcagaatttttagcagaacatgtacacaaaacttgaggggttgccttaaaaa 388
 ||||||||| ||||||||| ||||| ||||| ||||| |||||
 Db 177 CAGCTTCAGAAATCTTATGAGATATTCACAAAACCTTG--GGAGTAGTCTCTGAAAAA 234
 QY 389 ttgatgatctatatacaacatccgcttatt-aaaaatatctatcattacattacaa 447
 ||||| ||||||||| ||||| ||| ||||| ||||| ||||| |||||
 Db 235 CTAGATACCTCTATTCACATCCACTGTTAAAAAATAGATTCATTCCTCATTCAGAG 294
 QY 448 tgcgttttcagataatcagagaaatgccatcctcagaaatctgttaacaa 500
 ||||| ||||| ||||||||| ||||| ||| ||||| ||||| |||||
 Db 295 TGCTATTGTTGATGAATGAGAGAGATTCACAGAAAGTTGTATATCAAA 347
 RESULT 4
 BE008891/c 241 bp mRNA linear EST 05-JUN-2000
 LOCUS CM4-BN0161-040400-132-d08 BN0161 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BE008891
 ACCESSION BE008891
 VERSION BE008891.1 GI:8269124
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 241)
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 JOURNAL MEDLINE 20202653
 COMMENT Contact: Simpson A.J.C.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPSP/JCR Human Cancer Genome
 project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-CM4-BN0161-040
 400-132-008&t3=2000-04-04&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 6
 High quality sequence stop: 241.
 Location/Qualifiers
 1..241
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_1fb="BN0161"
 /dev_stage="Adult"
 /note="Organ: breast,normal; Vector: puc18; Site:1: Smal;
 Site:2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 BASE COUNT 73 a 40 c 47 g 81 t
 ORIGIN

Query Match 42.5%; Score 212.4; DB 9; Length 241;
 Best Local Similarity 99.5%; Pred. No. 1.4e-45;
 Matches 213; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 275 ctcattagactcagagatccatcgttagtcaatttagaagatgaatcgaagcagcctt 334
 |||||||
 Db 241 CTCATTAGACTCAGAGATCCATCGTTAGTCATTTTACCATGTAATCTGAAGCACACCTT 182

QY 335 ccgaatttttagcagacatgtfacaaaacttggagagtttccttaaaaaatgat 394
 |||||||
 Db 181 CCAGAAATTTTACGACATGTTGACAAAACCTGGAGGTTTGCTTAAAAAATTTGAT 122

QY 395 gcatctatacaacatccgccttaataaaatatatcatccacattaccaagtgcgtt 454
 |||||||
 Db 121 GCATCTATACACATCCGCTTATTAAAAATATATTTCATTCACCATTACCAAGTCCGCTT 62

QY 455 ttgcagataatggagaagatgcatcattgcagaat 488
 |||||||
 Db 61 TTGCAGATATATGGAAGAATGCATTCGACAAT 28

RESULT 5
 AZ644393 496 bp DNA linear GSS 14-DEC-2000
 LOCUS IM0508M18F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 DEFINITION clone UUGC1M0508M18 F, DNA sequence.
 ACCESSION AZ644393
 VERSION AZ644393.1 GI:11772878
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 496)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Rellily,
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.,
 and Wright, D., Weis, R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 CONTACT: Robert B. Weiss
 UNIVERSITY of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0508 row: M column: 18
 Seq primer: CGTTGTAACGACGCCAGT
 Class: plasmid ends
 High quality sequence stop: 496.
 Location/Qualifiers
 1. 496
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0508M18"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD42ny; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PMD42 (g147321149b/AF129072.1) a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptorized mouse DNA was annealed to
 adaptorized vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

Query Match 32.6%; Score 163.2; DB 12; Length 496;
 Best Local Similarity 86.5%; Pred. No. 1.6e-32;
 Matches 180; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 293 ccatcgttagcatctttagacgatgaactcgaagcagatccagatcttagcagac 352
 |||||
 Db 2 CCATCAGTACTGATTTAGATGATGAAGAACTGAACCTCCAGATCTTACGAT 61

QY 353 attgtacaaaacttggagggttgccttaaaaaatlagatgcatatatacaacatcg 412
 |||||||
 Db 62 ATTGTACAAAACCTTGAGGAGGATGTCCTGAAAAGACTATACCTTATTCAGCATCA 121

QY 413 ctattaaaaatatatcttaccatccatccaaagtcgttttgcagataaggagaag 472
 |||||||
 Db 122 CTGTGTAATAAATACATTCATTCCTCCACCTCCGAGTCTATTGTCAGATTAAGGAG 181

QY 473 atgcacatgcagaatgtgtaacaa 500
 |||||
 Db 182 ATACCTCTACAGAAGTGTGTAATCANA 209

RESULT 6
 BH110525 717 bp DNA linear GSS 19-JUL-2001
 LOCUS BH110525/c
 DEFINITION RPCI-24-340G10.TJ RPCI-24 Mus musculus genomic clone RPCI-24-340G10
 DNA sequence.
 ACCESSION BH110525
 VERSION BH110525.1 GI:14944731
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 717)
 Zhao, S., Nierman, W., Malek, J., Shatsman, S., Aklnret, B., Levins, M.,
 Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E.,
 Russell, D., de Jong, P., and Fraser, C.M.
 Mouse BAC End Sequences from Library RPCI-24
 Unpublished (1999)
 Other_GSSs: RPCI-24-340G10.TV
 CONTACT: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@tigr.org). Clones may be purchased from BACPAC
 Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
 page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Seq primer: SP6
 Plate: 340 row: G column: 10
 Class: BAC ends.
 Location/Qualifiers
 1. 717
 /organism="Mus musculus"

BASE COUNT	206 a	125 c	133 g	253 t
ORIGIN	/strain="C57BL/6J" /db_xref="taxon:10090" /clone="RPCI-24-340G10" /clone_id="RPCI-24" /sex="Male" /cell_type="Spleen/Brain" /note="Vector: pTRABAC1. Site_1: BamHI. Site_2: BamHI. pRCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTRABAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA." DNA.			
Query Match	31.8%; Score 159.2; DB 12; Length 717;			
Best Local Similarity	84.5%; Pred. No. 2e-31;			
Matches	191: Conservative	0: Mismatches	33: Indels	2: Gaps
OY	277	cattgaactcaagatccatcgtagtcattttagacgatgaatcgaacagacttc	336	
DB	717	CATCAGACTCAGAGATCCCATAGTACTATTTTAATGATGAAGAACTGACCTTCC	658	
OY	337	agaatttttagagacatttgcataaaacttggaggggtttgacctataaaattagatgc	396	
DB	657	AGAAATCTTAGAGATATTGTGACAAAACCTGGAGGGATGTGCTGGAAGACCTGGAATC	598	
OY	397	atctatacaacatccgctattataaaat--atctatctaccattaccagatgcgtt	454	
DB	597	CTCTATTGACGATCCACCTGTTAAAAATACACATCATTTCCCCACTCCGAGTGTATT	538	
OY	455	ttgcagataatggagaaagatgcgcatctgcagaatgtgttaatcaaa	500	
DB	537	TTGCAGATATATGGAGAAGATACCTCTACAGAAAGTTGTGAATCANA	492	
RESULT 7				
CNS04HOG/c	883 bp DNA linear GSS 21-MAY-2000			
LOCUS	CNS04HOG			
DEFINITION	Tetraodon nigroviridis genome survey sequence T7 end of clone 110L02 of library G from Tetraodon nigroviridis, genomic survey sequence.			
ACCESSION	AL291193			
VERSION	AL291193.1 GI:8029773			
KEYWORDS	GSS: genome survey sequence.			
SOURCE	Tetraodon nigroviridis.			
ORGANISM	Tetraodon nigroviridis			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorphae; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.			
AUTHORS	1 (bases 1 to 883) Roest-Crollius,H., Jallou,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.			
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 883) Roest-Crollius,H., Jallou,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F., Saurin,W. and Weissenbach,J.			
AUTHORS	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence			
TITLE	Unpublished			
JOURNAL	3 (bases 1 to 883)			
REFERENCE	Genoscope.			
AUTHORS	Direct Submission			
TITLE	Submitted (13-APR-2000) to the EMBL/Genbank/DBJ databases			
JOURNAL	This sequence is a single read and was generated as part of a large			
COMMENT	scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at			
	http://www.genoscope.cns.fr/tetraodon.			

FEATURES		Location/Qualifiers
Source		1..883
		/organism="Tetradodon nigroviridis"
		/db_xref="taxon:99883"
		/clone="110L02"
		/clone_lib="c"
		/note="Genoscope sequence ID : COBG110D01LPI-end : T7"
BASE COUNT	233 a 227 c 197 g 223 t	3 others
ORIGIN		
Query Match	28.4%	Score 142; DB 12; Length 883;
Best Local Similarity	59.8%	Pred. No. 7e-27;
Matches 238; Conservative	0; Mismatches 160; Indels	0; Gaps 0;
OY	8	acagaagaacacatctaccagctcagctgcagctttaaaccagaacgatttgcacgtcttacc 67
DB	398	ACAGGGAACACTTGTACTCAGCTGCAATGCTGAGCCCGGCGGCTCGACTCATC 339
OY	68	aaggaagtaatgaatcacatcttcgctgcgcagagaaatgattgttcaatggtatccatt 127
DB	338	AAGGACGCTGCTCTCTCAGCTGAGCCCGGAGGAGGACTTCGCTGTGGAATGGAGACCTGGA 279
OY	128	gatgaacaacagaaatccaccatctggtttcaatggttcaatggttggaaaaatccttat 187
DB	278	AACAGAGAGCTGAGACATCCACCCCTTTCCTGCTCAGAGATGCTTGGAMACATCTCTAC 219
OY	188	atacatttttcagagatttgaatttattttatgtatgatgcacattatccccaagacata 247
DB	218	ATACATTTTGTCTGAGGAGCTGAGCACTTTGCANAGACATGCTTTGATGCCACTGCTGCA 159
OY	248	ctagaagaaggtgcagacatgctgtggaactcatatgaactcaggaattccatcgctgaatc 307
DB	158	CTTGAGGAAACATGACGAGATTTCATCTCTGCGGCTCAGACACTCTTCAACCATCATTA 99
OY	308	ttagaagatgaatcttgaagcacaagcttccagaatttttaagcagacatgtgtcaaaaact 367
DB	98	TTTGGGATGCGGAGAAAGACGACACCCCTCTGGAAACCTTTCGAGATCATGTGAAAGCTC 39
OY	368	ggaaggttgcctcttaaaaaattgaatgacatctata 405
DB	38	GGAGGGACGGTCATGACAAAGCTGGATTCGTGTTGCA 1
RESULT 8		
LOCUS	BB561400	
DEFINITION	BB561400 RIKEN full-length enriched, 10 days neonate olfactory	
ACCESSION	BB561400	
VERSION	BB561400.1	
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Enkayotla; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS	1 (bases 1 to 883)	
	Komno, H., Aitawa, K., Akahira, S., Akiyama, J., Arahawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hata, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, T., Itoh, M., Iizawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, T., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomihata, N., Toyama, T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamana, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.	
TITLE	RIKEN Mouse ESTs (komno, H., et al.)	
JOURNAL	Unpublished (2000)	
COMMENT	Contact: yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic	

JOURNAL Unpublished (2001)
 COMMENT Other_GSSs: BOHLS34TF
 Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdlowne@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TR
 Class: sheared ends.

FEATURES
 Source
 1..752
 /organism="Brassica oleracea"
 /strain="T01000Dh3"
 /db_xref="taxon:3712"
 /clone="BOHLS54"
 /clone_lib="BOHL"
 /note="Vector: pHOSt; Site_1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOSt using BstXI linkers"

BASE COUNT 267 a 114 c 109 g 262 t
 ORIGIN

Query Match 8.8%; Score 44; DB 12; Length 752;
 Best Local Similarity 45.1%; Pred. No. 0.41;
 Matches 164; Conservative 0; Mismatches 200; Indels 0; Gaps 0;

QY 124 attgataaacaacgaataccaccatcgttcctcatggtcctaagaatggttggaaaact 183
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 466 ATGACATGGCTACGAAATTCATGACATGCTTTATAGCTTATATGACATGTACATTGC 407
 QY 184 ttatacatctttcagaagatttbaactttatgatgagatccactatccacagac 243
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 406 ATTATTATGTATTTATTTTGGTGAACCTTTTAAATATGGAATATATTCATATATT 347
 QY 244 tatactagaagagagtcacagatgltgaaactatagactcagatccatcgtagt 303
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 346 ACATTATATGTCATTTATTTTGGAAATTTTATGAAATGTCAACTATAATCAT 287
 QY 304 catttgaagcagatctgaagcagctccagaatttttagcagacatgtgcaaaa 363
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 286 CATTAAATTAATATATTCAAATATGACATATTAATTCGAAATATTAATTAAT 227
 QY 364 acttgagaggttctccctaaataatgaatgacatctatcaaacctcgctatataaaa 423
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 226 ATTTTAAATTAATACATTTTATATACATAAAATTTTCAAAAATGCAATACATTTT 167
 QY 424 atatacatctcacatctacacagtgctgttttcagataaalgaggaagatgcattgca 483
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 166 AAATTAATTAATTTAATGCTAAATCATTAATTTTCTCTGTTGACGGCGGACGATTAAA 107
 QY 484 gaaa 487
 |||||
 Db 106 GAAA 103

RESULT 12
 A2066617/c
 LOCUS
 DEFINITION RPCI-23-333L3.TV RPCI-23 Mus musculus genomic clone RPCI-23-433L3,
 DNA sequence.
 ACCESSION A2066617
 VERSION A2066617.1 GI:7357869
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 440)
 Zhao, S., Nierman, W., Feldblum, T., Malek, J., Shatsman, S., Akhmet
 B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Kroll, M., de Jong, P.,
 and Fraser, C.M.

TITLE Mouse BAC End Sequences from library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Other GSSs: RPCI-23-433L3.TJ

Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@edlong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 433 row: L column: 3
Seq primer: T7
Class: BAC ends.

FEATURES
source Location/Qualifiers

1..440
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-433L3"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 120 a 74 c 73 g 173 t
ORIGIN

Query Match 8.8%; Score 43.8; DB 12; Length 440;
Best Local Similarity 87.3%; Pred. No. 0.4;
Matches 48; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 446 agtgcgtcttgcagataatgagaaagtcacatgcagaaatgtgtaacaa 500
|||||
Db 418 AGTGATATTTTGCAGATATGAGAGATACCTCTACAGAGTGTGTATCAAA 364

RESULT 13
BH131081 895 bp DNA linear GSS 07-AUG-2001
LOCUS ENTROP16TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic DNA sequence.
ACCESSION BH131081
VERSION BH131081.1 GI:15089550
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 895)
AUTHORS Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library (2001)
COMMENT Unpublished (2001)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library

Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 43
High quality sequence stop: 830.
Location/Qualifiers

FEATURES
source

1..895
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: PHOS1; Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a light size distribution (~2 kb). The v + 1 method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 302 a 117 c 103 g 373 t
ORIGIN

Query Match 8.3%; Score 41.4; DB 12; Length 895;
Best Local Similarity 50.2%; Pred. No. 2.1; Length 895;

Matches 102; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

Qy 32 cagcttciaatcacagaacgattgcacgctctatacagaagtaatgaatcatctg 91
|||
Db 324 CAATTTGTTAACAAAGACAATTTGATGAGAAACAAGCAAGAAAGAGAAATTTTAT 265
Qy 92 cctgcagagaaattgattcaatggtatccatttgatgaacacgaacacccatc 151
|||
Db 264 ACTTCATTAGTACTGTTGTTATTTATTCCTTACCTTAATTCAGAAAGAAATCTGAATTA 205
Qy 152 gtttcagtcgaatgagtggttggaacacatctatatacatcttcagagagttgact 211
|||
Db 204 TTTCAAAATATCTAATATTCCTTTTAAACAGATTATTTTAAATTTCTACAGAACATTA 145
Qy 212 ttattgatgagatgcacattat 234
|||
Db 144 ACAATTAATTAATAAGTTACTTAT 122

RESULT 14
BH136596 896 bp DNA linear GSS 07-AUG-2001
LOCUS ENTMC65TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic DNA sequence.
ACCESSION BH136596
VERSION BH136596.1 GI:15095657
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 896)
AUTHORS Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library (2001)
COMMENT Unpublished (2001)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library

Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 47
High quality sequence stop: 723.
Location/Qualifiers
1. 896

FEATURES

source

/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: PHOS1; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT
ORIGIN

402 a 70 c 130 g 294 t

Query Match 8.3%; Score 41.4; DB 12; Length 896;
Best Local Similarity 50.2%; Pred. No. 2.1;
Matches 102; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 32 cagcttctaataccagaagatttgcagcgtctatcaagaagaatgaatcattctcg 91
DB 154 CAATTTGTTAAACAAGACAATGATGAGAAACAAGTAAGAAGAATTTTAT 213
QY 92 cctggcagagaattgattgtaaatgtaacattgataaacaagaatcaccatct 151
DB 214 ACTTCATTACTAGTATTGTTATTATCTTCACCTAATTCAAGGAATGCTGAATTA 273
QY 152 gtttcattgcttaagaatggttggaaaatccttatatacatcttccaaggattgact 211
DB 274 TTACAAAAATCTAAATCCCTTTTAAACAGTATTAAATAATTCYAGAAGACATACT 333
QY 212 ttattgatgagatgccactat 234
DB 334 ACAATTATATAAAGTTACTTAT 356

RESULT 15

A2547902

LOCUS 940 bp DNA linear GSS 14-NOV-2000

DEFINITION ENTDP1ITF Entamoeba histolytica Sheared DNA Entamoeba histolytica

genomic, DNA sequence.

ACCESSION

A2547902

VERSION

A2547902.1 GI:11170987

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library

Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 299
High quality sequence stop: 873.
Location/Qualifiers
1. 940

FEATURES

source

/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: PHOS1; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT
ORIGIN

393 a 92 c 149 g 306 t

Query Match 8.3%; Score 41.4; DB 12; Length 940;
Best Local Similarity 50.2%; Pred. No. 2.1;
Matches 102; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 32 cagcttctaataccagaagatttgcagcgtctatcaagaagaatgaatcattctcg 91
DB 483 CAATTTGTTAAACAAGACAATGATGAGAAACAAGTAAGAAGAATTTTAT 542
QY 92 cctggcagagaattgattgtaaatgtaacattgataaacaagaatcaccatct 151
DB 543 ACTTCATTACTAGTATTGTTATTATCTTCACCTAATTCAAGGAATGCTGAATTA 602
QY 152 gtttcattgcttaagaatggttggaaaatccttatatacatcttccaaggattgact 211
DB 603 TTACAAAAATCTAAATCCCTTTTAAACAGTATTAAATAATTCYAGAAGACATACT 662
QY 212 ttattgatgagatgccactat 234
DB 663 ACAATTATATAAAGTTACTTAT 685

Search completed: May 22, 2002, 05:31:11
Job time: 4071 sec

.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 06:59:14 ; Search time 3530.57 Seconds
(without alignments)
118.545 Million cell updates/sec

Title: US-09-693-205-47

Perfect score: 20
Sequence: 1 gtagatgcacattgtcact 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vl:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*
29: em.vl:*
30: em.htg.hum:*
31: em.htg.inv:*
32: em.htg.other:*
33: em.htgo.inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query No. Score Match Length DB ID Description

1	20	100.0	12793	6	AX119931	AX119931 Sequence
2	20	100.0	12793	9	AF193556	AF193556 Homo sapi
3	20	100.0	92693	9	AL157766	AL157766 Human DNA
4	20	100.0	99819	2	AC079761	AC079761 Homo sapi
5	18.4	92.0	11492	6	AX119933	AX119933 Sequence
6	18.4	92.0	11493	10	AF193557	AF193557 Mus muscu
7	17.4	87.0	434	11	G37224	G37224 SHGC-57260
8	17.4	87.0	766	3	AF113358	AF113358 Ips latid
9	17.4	87.0	766	3	AF113359	AF113359 Ips latid
10	17.4	87.0	175309	2	AC069021	AC069021 Homo sapi
11	17.4	87.0	176713	9	AP000751	AP000751 Homo sapi
12	17.4	87.0	192913	2	AP001265	AP001265 Homo sapi
13	17.4	87.0	216521	9	AC073587	AC073587 Homo sapi
14	17.4	87.0	241392	2	AC019279	AC019279 Homo sapi
15	17	85.0	993	3	P095073	P095073
16	17	85.0	8028	3	MIPOND2	MIPOND2
17	17	85.0	69648	2	AC099825	AC099825 Papio ham
18	17	85.0	129504	9	AL445469	AL445469 Human DNA
19	17	85.0	160866	2	AC092532	AC092532 Papio cyn
20	17	85.0	174897	2	AC090965	AC090965 Papio cyn
21	17	85.0	178317	9	AC012081	AC012081 Homo sapi
22	17	85.0	201957	2	AC026359	AC026359 Homo sapi
23	16.8	84.0	2544	2	AC020530	AC020530 Drosophi1
24	16.8	84.0	31887	9	AL450999	AL450999 Human DNA
25	16.8	84.0	61394	2	AC102771	AC102771 Mus muscu
26	16.8	84.0	68997	2	DMB6F4	DMB6F4
27	16.8	84.0	71999	2	AC101210	AC101210 Mus muscu
28	16.8	84.0	72648	2	AC101171	AC101171 Mus muscu
29	16.8	84.0	72648	2	AC101171	AC101171 Mus muscu
30	16.8	84.0	84894	8	AP003251	AP003251 Oryza sat
31	16.8	84.0	118380	2	AC094213	AC094213 Rattus no
32	16.8	84.0	123943	10	AL592547	AL592547 Mouse DNA
33	16.8	84.0	130084	2	AC092265	AC092265 Homo sapi
34	16.8	84.0	152042	9	AL137857	AL137857 Human DNA
35	16.8	84.0	166704	2	AC078821	AC078821 Homo sapi
36	16.8	84.0	169576	2	AL663043	AL663043 Mus muscu
37	16.8	84.0	170669	2	AC007907	AC007907 Homo sapi
38	16.8	84.0	171366	2	AC079465	AC079465 Homo sapi
39	16.8	84.0	173525	2	AC067893	AC067893 Homo sapi
40	16.8	84.0	182004	2	AC097609	AC097609 Rattus no
41	16.8	84.0	184649	2	AC025185	AC025185 Homo sapi
42	16.8	84.0	188553	2	AC093394	AC093394 Bos tauru
43	16.8	84.0	210730	2	AC015975	AC015975 Homo sapi
44	16.8	84.0	216016	2	AL626775	AL626775 Mus muscu
45	16.8	84.0	228550	2	AC093396	AC093396 Bos tauru

ALIGNMENTS

RESULT	1					
LOCUS	AX119931	AX119931	12793 bp	DNA	linear	PAT 11-MAY-2001
DEFINITION	Sequence 1 from Patent WO0129266.					
ACCESSION	AX119931					
VERSION	AX119931.1	GI:14036678				
KEYWORDS						
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
TITLE	Hudson,T.J., Engert,J. and Richter,A.					
JOURNAL	Identification of arsacs mutations and methods of use therefor					
FEATURES	Patent: WO 0129266-A 1 26-APR-2001;					
source	McGILL UNIVERSITY (CA) : Hospital Sainte-Justine (CA)					
	Location/Qualifiers					
	1..12793					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
	BASE COUNT	4163 a	2256 c	2487 g	3887 t	
	ORIGIN					

AUTHORS Tromans, A.
TITLE Direct Submission
JOURNAL Submitted (11-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquyer@sanger.ac.uk

COMMENT

Requests: clonequest@sanger.ac.uk
On April 12, 2001 this sequence version replaced g1:12709868.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw, S, SWISSPROT; Tr, TrEMBL; Wp, WormPep; Information on the WormPep
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr13>
RP11-40020 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
<http://www.chori.org/dacpac/home.htm>
VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone.
RP11-40020 it may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP11-760M1 is at 92594 in this sequence.
The true right end of clone RP11-72P19 is at 100 in this sequence.

FEATURES

source

1..92693
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="g12.11-12.2"
/clone="RP11-40020"
/clone_1lb="RPCI-11.1"
2390..2485
/note="MIR repeat: matches 81..192 of consensus"
2562..2673
/note="MIR repeat: matches 68..183 of consensus"
3896..4201
/note="AluY repeat: matches 3..308 of consensus"
5122..5397
/note="MER46C repeat: matches 1..286 of consensus"
18986..19294
/note="AluY repeat: matches 1..310 of consensus"
19644..19873
/note="MER46A repeat: matches 1..235 of consensus"
20613..20912
/note="AluSg1 repeat: matches 1..300 of consensus"
23342..23651
/note="AluSg1 repeat: matches 1..309 of consensus"
24769..24891
/note="L2 repeat: matches 2554..2682 of consensus"
23871..26011
/note="L2 repeat: matches 2356..2495 of consensus"
26033..26109
/note="L2 repeat: matches 2601..2688 of consensus"
26245..26344
/note="L2 repeat: matches 2154..2255 of consensus"
26338..27096
/note="MIR repeat: matches 3..175 of consensus"
27150..27653
/note="L2 repeat: matches 1063..1644 of consensus"

repeat_region 28522..28891
/note="TMR1B repeat: matches 1..364 of consensus"
repeat_region 29447..29834
/note="LIME3A repeat: matches 5787..6164 of consensus"
repeat_region 36098..36415
/note="AluSx repeat: matches 1..308 of consensus"
repeat_region 37202..37414
/note="MIR repeat: matches 22..262 of consensus"
repeat_region 37963..38254
/note="AluSg repeat: matches 9..301 of consensus"
repeat_region 38703..39008
/note="AluSg repeat: matches 1..306 of consensus"
repeat_region 39790..40093
/note="AluSx repeat: matches 1..304 of consensus"
repeat_region 40126..40416
/note="AluSg repeat: matches 1..292 of consensus"
repeat_region 40444..40733
/note="AluSg repeat: matches 1..292 of consensus"
misc_feature 41322..41405
/note="Single clone region. Assembly confirmed by
restriction digest data."
repeat_region 41541..41788
/note="AluSg repeat: matches 1..248 of consensus"
repeat_region 44790..45101
/note="AluSg repeat: matches 1..313 of consensus"
repeat_region 45261..45312
/note="13 copies 4 mer tgtt 888 conserved"
repeat_region 45899..46206
/note="AluY repeat: matches 1..307 of consensus"
repeat_region 46754..47052
/note="AluY repeat: matches 1..298 of consensus"
repeat_region 47067..47365
/note="AluY repeat: matches 1..299 of consensus"
repeat_region 47477..47873
/note="L1MA10 repeat: matches 5950..6322 of consensus"
repeat_region 47889..48229
/note="AluSx repeat: matches 1..312 of consensus"
repeat_region 49168..49212
/note="Alu repeat: matches 85..126 of consensus"
repeat_region 49620..49693
/note="L2 repeat: matches 1685..1757 of consensus"
repeat_region 50704..51032
/note="AluSx repeat: matches 1..308 of consensus"
misc_feature 52204..53009
/note="CpG Island"
repeat_region 53978..54137
/note="AluSx repeat: matches 1..160 of consensus"
repeat_region 54179..54511
/note="L1MB6 repeat: matches 5822..6172 of consensus"
repeat_region 55685..55949
/note="AluSx repeat: matches 1..305 of consensus"
repeat_region 57331..57390
/note="30 copies 2 mer ga 75% conserved"
repeat_region 57357..57392
/note="9 copies 4 mer gaga 91% conserved"
repeat_region 57753..57930
/note="MIR repeat: matches 82..262 of consensus"
repeat_region 58260..58389
/note="MIR repeat: matches 2..153 of consensus"
repeat_region 58564..58611
/note="24 copies 2 mer ca 93% conserved"
repeat_region 59350..59533
/note="AluSg repeat: matches 129..313 of consensus"
repeat_region 59992..60223
/note="AluSg repeat: matches 85..299 of consensus"
repeat_region 61036..61144
/note="L2 repeat: matches 2581..2696 of consensus"
repeat_region 62008..62187
/note="TIGER1 repeat: matches 2238..2418 of consensus"
repeat_region 62188..62316
/note="AluY repeat: matches 1..129 of consensus"
repeat_region 62330..62363

```
repeat_region /note="Alu repeat: matches 261. .294 of consensus"
62362. .62565
/note="TIGER1 repeat: matches 1586. .1787 of consensus"
repeat_region 62566. 62865
/note="AluSg repeat: matches 1. .302 of consensus"
repeat_region 62866. 64385
/note="TIGER1 repeat: matches 46. .1586 of consensus"
repeat_region 64386. 64694
/note="AluY repeat: matches 1. .306 of consensus"
repeat_region 64695. 64713
/note="TIGER1 repeat: matches 29. .46 of consensus"
repeat_region 65068. 65395
/note="L1PB2 repeat: matches 5405. .5733 of consensus"
repeat_region 65396. 65569
/note="AluY repeat: matches 136. .309 of consensus"
repeat_region 65571. 65640
/note="L1PB2 repeat: matches 5728. .5791 of consensus"
repeat_region 65696. 65717
/note="11 copies 2 mer ta 100% conserved"
repeat_region 65725. 66096
/note="L1PB2 repeat: matches 5789. .6155 of consensus"
repeat_region 66371. 66410
/note="10 copies 4 mer tctg 82% conserved"
repeat_region 67586. 67886
/note="AluY repeat: matches 1. .299 of consensus"
repeat_region 69748. 69930
/note="MIR repeat: matches 6. .248 of consensus"
repeat_region 70957. 71267
/note="AluY repeat: matches 1. .311 of consensus"
repeat_region 71279. 71413
/note="MER21B repeat: matches 548. .680 of consensus"
repeat_region 71411. 71737
/note="MER31A repeat: matches 47. .485 of consensus"
repeat_region 71780. 72075
/note="AluSx repeat: matches 1. .295 of consensus"
repeat_region 72143. 72236
/note="MER31-internal repeat: matches 42. .175 of
consensus"
repeat_region 72454. 72865
/note="MER31-internal repeat: matches 332. .739 of
consensus"
repeat_region 72873. 73249
/note="MER31-internal repeat: matches 883. .1261 of
consensus"

Query Match 100.0%: Score 20; DB 9; Length 92693;
Best Local Similarity 100.0%: Pred. No. 3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtaatggcacttgcact 20
Db 12437 GTGATGCGCACCTTGCAC 12418
```

```
RESULT 4
AC079761 99819 bp DNA linear HTG 10-SEP-2000
LOCUS Homo sapiens chromosome UNK clone RP11-143G17, *** SEQUENCING IN
DEFINITION
AC079761 AC079761.1 GI:10047966
VERSION AC079761.1
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 99819)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 99819)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL 4 Submitted (10-SEP-2000) Genome Sequencing Center, Washington
```

COMMENT

```
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 44 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1137: contig of 1137 bp in length
1138 1237: gap of unknown length
1238 2538: contig of 1301 bp in length
2539 2638: gap of unknown length
2639 3976: contig of 1338 bp in length
3977 4076: gap of unknown length
4077 5355: contig of 1279 bp in length
5356 5455: gap of unknown length
5456 6757: contig of 1302 bp in length
6758 6858: gap of unknown length
6859 8570: contig of 1713 bp in length
8571 8671: gap of unknown length
8672 9925: contig of 1255 bp in length
9926 10025: gap of unknown length
10026 11426: contig of 1401 bp in length
11427 11526: gap of unknown length
11527 13266: contig of 1740 bp in length
13267 13366: gap of unknown length
13367 14794: contig of 1428 bp in length
14795 14894: gap of unknown length
14895 16054: contig of 1160 bp in length
16055 16154: gap of unknown length
16155 17395: contig of 1241 bp in length
17396 17495: gap of unknown length
17496 19287: contig of 1792 bp in length
19288 19387: gap of unknown length
19388 21294: contig of 1907 bp in length
21295 21394: gap of unknown length
21395 22944: contig of 1550 bp in length
22945 23044: gap of unknown length
23045 24421: contig of 1377 bp in length
24422 24521: gap of unknown length
24522 25870: contig of 1349 bp in length
25871 25970: gap of unknown length
25971 27230: contig of 1260 bp in length
27231 27330: gap of unknown length
27331 28778: contig of 1448 bp in length
28779 28878: gap of unknown length
28879 30893: contig of 2015 bp in length
30894 30993: gap of unknown length
30994 32460: contig of 1467 bp in length
32461 32560: gap of unknown length
32561 33984: contig of 1424 bp in length
33985 34084: gap of unknown length
34085 35285: contig of 1201 bp in length
35286 37184: contig of 1799 bp in length
37185 37284: gap of unknown length
37285 39172: contig of 1888 bp in length
39173 39272: gap of unknown length
39273 40874: contig of 1602 bp in length
40875 42893: contig of 1919 bp in length
42894 42993: gap of unknown length
42994 44384: contig of 1391 bp in length
44385 44484: gap of unknown length
44485 45999: contig of 1515 bp in length
```



```
* 46000 46099: gap of unknown length
* 46100 48669: contig of 2570 bp in length
* 48670 48670 48669: gap of unknown length
* 48770 50798: contig of 2029 bp in length
* 50799 50898: gap of unknown length
* 50899 52809: contig of 1911 bp in length
* 52810 52909: gap of unknown length
* 52910 55127: contig of 2218 bp in length
* 55228 55227: gap of unknown length
* 55228 58087: contig of 2860 bp in length
* 58088 58187: gap of unknown length
* 58188 61004: contig of 2817 bp in length
* 61005 61104: gap of unknown length
* 61105 64185: contig of 3081 bp in length
* 64186 64285: gap of unknown length
* 64286 67105: contig of 2820 bp in length
* 67106 67205: gap of unknown length
* 67206 70837: contig of 3632 bp in length
* 70838 70937: gap of unknown length
* 70938 75837: contig of 4900 bp in length
* 75838 75937: gap of unknown length
* 75938 80452: contig of 4515 bp in length
* 80453 80552: gap of unknown length
* 80553 84661: contig of 4109 bp in length
* 84662 84761: gap of unknown length
* 84762 90542: contig of 5781 bp in length
* 90543 90642: gap of unknown length
* 90643 94348: contig of 3706 bp in length
* 94349 94449: gap of unknown length
* 94449 99819: contig of 5371 bp in length.
```

FEATURES
source

```
1. 99819  
/organism="Homo sapiens"  
/db_xref="taxon:9606"  
/chromosome="UNK"  
/clone="RP11-143617"  
1. 1137  
/note="assembly_name:Contig20"  
misc_feature 1238..2538  
/note="assembly_name:Contig23"  
misc_feature 2639..3976  
/note="assembly_name:Contig30"  
misc_feature 4077..5355  
/note="assembly_name:Contig35"  
misc_feature 5456..6757  
/note="assembly_name:Contig39"  
misc_feature 6858..8570  
/note="assembly_name:Contig40"  
misc_feature 8671..9925  
/note="assembly_name:Contig41"  
misc_feature 10026..11426  
/note="assembly_name:Contig42"  
misc_feature 11527..13266  
/note="assembly_name:Contig44"  
misc_feature 13367..14794  
/note="assembly_name:Contig47"  
misc_feature 14895..16054  
/note="assembly_name:Contig48"  
misc_feature 16155..17395  
/note="assembly_name:Contig49"  
misc_feature 17496..19287  
/note="assembly_name:Contig50"  
misc_feature 19388..21294  
/note="assembly_name:Contig51"  
misc_feature 21395..22944  
/note="assembly_name:Contig52"  
misc_feature 23045..24421  
/note="assembly_name:Contig53"  
misc_feature 24522..25870  
/note="assembly_name:Contig54"  
misc_feature 25971..27230  
/note="assembly_name:Contig55"  
misc_feature 27331..28778
```

```
misc_feature /note="assembly_name:Contig57"  
28879..30893  
/note="assembly_name:Contig58"  
misc_feature 30994..32460  
/note="assembly_name:Contig59"  
misc_feature 32561..33984  
/note="assembly_name:Contig60"  
misc_feature 34085..35285  
/note="assembly_name:Contig61"  
misc_feature 35386..37184  
/note="assembly_name:Contig62"  
misc_feature 37285..39172  
/note="assembly_name:Contig63"  
misc_feature 39273..40874  
/note="assembly_name:Contig64"  
misc_feature 40975..42893  
/note="assembly_name:Contig65"  
misc_feature 42994..44384  
/note="assembly_name:Contig66"  
misc_feature 44485..45999  
/note="assembly_name:Contig67"  
misc_feature 46100..48669  
/note="assembly_name:Contig68"  
misc_feature 48770..50798  
/note="assembly_name:Contig69"  
misc_feature 50899..52809  
/note="assembly_name:Contig70"  
misc_feature 52910..55127  
/note="assembly_name:Contig71"  
misc_feature 55228..58087  
/note="assembly_name:Contig72"  
misc_feature 58188..61004  
/note="assembly_name:Contig73"  
misc_feature 61105..64185  
/note="assembly_name:Contig74"  
misc_feature 64286..67105
```

Query Match 100.0%; Score 20; DB 2; Length 99819;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gtgaatggccacttgacct 20
Db 87022 GTGAATGGCCACTTTCACCT 87003
|||||

RESULT 5
AX119933 11492 bp DNA Linear PAT 11-MAY-2001
LOCUS
DEFINITION Sequence 3 from Patent W00129266.
ACCESSION AX119933
VERSION AX119933.1 GI:14036679
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
1 (bases 1 to 11492)
AUTHORS Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.
TITLE Identification of arcs mutations and methods of use therefor
JOURNAL Patent: WO 0129266-A 3 26-APR-2001;
MCGILL UNIVERSITY (CA); Hopital Sainte-Justine (CA)

FEATURES
source 1. 11492
/organism="Mus musculus"
/db_xref="taxon:10090"
BASE COUNT 3599 a 2280 c 2387 g 3226 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 6; Length 11492;
Best Local Similarity 95.0%; Pred. No. 22;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Oy      1  gtaaatggccacttgact 20
        |||
        6396 GTGAATGGCCACTTGTCTCT 6415

RESULT  6
AF193557 11493 bp DNA linear ROD 07-FEB-2000
LOCUS Mus musculus sacsin gene, complete cds.
DEFINITION AF193557
ACCESSION AF193557.1 GI:6907043
VERSION
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 11493)
AUTHORS Engert,J.C., Berube,P., Mercier,J., Dore,C., Lepage,P., Ge,B.,
Bouchard,J.P., Mathieu,J., Melancon,S.B., Schalling,M.,
Lander,E.S., Morgan,K., Hudson,T.J. and Richter,A.
ARSCS) a spastic ataxia common in northeastern Quebec, is caused
by mutations in a new gene encoding an 11.5-kb ORF
20120709
2 (bases 1 to 11493)
Engert,J.C., Berube,P., Dore,C., Lepage,P., Ge,B., Hudson,T.J. and
Richter,A.
Direct Submission
Submitted (08-Oct-1999) Genome Centre, Montreal General Hospital,
1650 Cedar Ave., Montreal, QC H3G 1A4, Canada
Location/Qualifiers
1. 11493
/organism="Mus musculus"
/db_xref="taxon:10090"
<1. >11493
/product="sacsin"
1. 11493
/note="molecular chaperone"
/codon_start=1
/product="sacsin"
/protein_id="AAF31263.1"
/db_xref="GI:6907043"
/translation="MNTFPGRELIVVOMYFSEDRKHPSLSWLKMVKNLYTHFSEDL
TLFDEMLPRLINEDOTCEVILRLIPSVYIILDEPFAOLPEPLADIVOKIGLIVL
KRLDSIOHPLVKKYIHSPLSALIIQIMKRIPLQKCNDAISALIPHRDAKPLASL
TDISEKRIIOELTIFKRIINSSDQISSYKLGCKVLDFTAKLPDRLRSVSD
SSDEATIRLANMLIEKLTCTCLFVLKDIGNAEYTODEVQMLMILENLSLKE
NSNVLDWMLPLKFIHMSOGHVAAGDLDPDIEVDLDFEENEACFPPTIPSPDL
HSLRQIGLNESLSKERDVGVAARKTEALVQSSCONODVLMKAKTLLLVKNQTL
OSSEGMALKIKVYPACKERPNTPGSLVMGDLINCAAPPDMCDAHAVLVGSSTP
LVESYVNLLEQALSIPTKRTIYAVLKHRTTYVDYTSKTFSEDIYQPHIILEYGF
MHDLISEGDSFKALFPMWWTGKNFCPLAQAIVIPTHDLIDQPLVLYNPQMAKPHQ
LEKAGSIEELTSDHISWYIQVYKSDLESSESKOMLMLIMRLVYNQIPAS
PNTPVPIYHSRNPSTLYMKPIHECCYCDIKVDLNDLEDSEYIILVEDIPMTAE
WTKVPCLSRLINPENMGFEQSGREPLVRIKNTLEEPSYDIKFELQADANANA
TSCSMIMRRMNDIRENLDPGMAAGHPCGPHIGEVGCLPLRIKIGLPIHNGCAVNSR
EYDKGKFCIGNSVYHTDPIIMASREPMIMPDNINISHIKIDRSPGKINMSK
OQKRLKFPNOKRPIDVFCOLPLAVEAPVSYNGFLRSLSTRQDEAVSSEVSTCY
NTADIVSLVDEFSICGHRILITFOSVNSVYLYKLIETNPISADDTIIRKRVCEP
ALNAVILSVLEAKALMKTCSSNNKLPIDVYKSSCIQIIEVEHFHVRADLQSP
LERGDDDPATLEFMAKSGOSKPSDELPOKVDCTWMLICMDTEGALKRSLNSG
BRGLVPGGAVGLVLEHTOEOKWTVKPHIGEVGCLPLRIKIGLPIHNGCAVNSR
KEIKTDPKGRNNTFMNRVITYKALQALSVRLDALIGELIDTYTYAAMPDLYHD
DESVICGFTEDIAHGKGEILTRFSDGSWMVSMKNVRLLDSIIQRKVGSAARFIE
LKLYKTGSKNCAVELPSVKAEGEACQOILENTEFSEQFSEVFEPNIOETEA
ELRDLPMNVLEKLEDEFSGLIRTVCPVCSLEGHPLVPSRLIPEGVARLFTKD
GREPYGODVYLNPIILIKLVOLGAKDILMDMLERESVAEINKSDHAACLRSS
ILSLIDEKTIKIDPRAKDFAKQTIPTPLPTKAGSLKMGKSPFEMPAATD
ITYAEODIVCLLOPLINENSGFRGCGSVSLAVKEFGLKRPYDVLINQKQVAK
SVDDGITLQENTNACIKYILHEAVLQNMARATIIIEKIPCFIILVENVYSEKVS
FLNFEAAVYLVQLPNPKRNNRELFEVSGRQSTVEDFALVLSIDDERKQIOTE
ENFOLCRRIISGIMSLIREKQEFCEKNYKILLPDTNLLLPKSLCYNDCPMIKY

```

```

KDSIVKCHADIPREAVAKGAIPIKRNKALERYASNICEFALTEFGEGEKLTSRIS
ILNAVSEKEMKELLONADAKATEICVPPRODHPVIRIDFMDPLOGALCYN
NOPETEDVRGIONKGTCKBGNPKCTGYIGIFNSVYHITDPCPSINDLIGIEDP
HARIAPGATSVSPGRMRJDLUDFRTORSVDYLDILGNHFRKNDNTMRPREPLRNEMA
OVSEISVSPSSDRMVONLIDKRLRSGAELMTFNMHEKISTICEIDAKTGLVNLVSVK
QKITDGDRLRKROFHASVYDTSYTKRKROLKIDIPVQOITYMDDESGNLTJMLICNRS
GFSMEKYSKSVISAHKNODITLPPRGVAACITNHNKPHAPFELPLSTETGLPFI
VNGHFDALDSARRNLWRDNGVGRSDMNNSTLIALPARYVELLIOLKRRYPGSDPT
LSVIONTPHIVYITIKRFLSEFPVNRNDIDOPDLCIVKALYSCHHEMDKRLPVA
PNDSDSLHSAVITITWINKSTNRTPPFDNLQDEPLKADNNTITRKVAENVY
RUKHLLLETGFNLVNCDEFTANLYCHLVADNIPVSTVYPADVRSFLMFTSPDNCHI
GKLPCLRQOTNLKLFHSLKLVYDCEKAESESEFEVEGLPLITLDSVLIQIDGRKPK
FLTYHELIPSRKDFMNTLYLKYSVLLNCAVAKVAFVSDISFADLSYLPREYTKN
CAKMDNFASBSMLKNAHFISESIVTDQDEPPAPVYIDILKMDALLPGETFY
SNOLVVPBGDVLIPLSLMHIAPFPNAQSDKFKHLMAGACIQIOLNKCSDSLVP
LUSCHTANIDSASILKAVHYVQSTRTFTEKLEMDPEBALMTNCSLHMSDDI
KLKLSIPCKYSISGRMSTAKFGTCYVLKSPSAVEVEMWTOSSAFIEEKVLEKEL
YVLGCPVDDLEVLYLKLPLKTEINLSDAKLEHLIYLNRLASIEEPSEIKOELEK
LESLLIHDANRLKQAKHFYDRTVAFVEMLPEKLIPEKFEKLEQVIKRNQAF
MNSWEEFLNIGLKALSOOOLQFAKEISVARNENMSKETIOTSDVIDLHIFQER
MOLISGNFLKELSLIPLCPERRAPAVYIRFHPYOEVNCTIPLIKFNAQVNRKRCOC
DYLOLMTISCPILPEKATPLSTKEQDSLRAQEOLEOYLNMLNPNVDPRLDYVINC
RNICNTITLDEEMVTRAKVLRSTIYEFUSAERRERFQURGVAFVWEDGMRLPEE
VYINLEYEADFPYLYKLELEGTGFHQLEKLGTDIISTKOYVAVLSEKSSGQ
LDPNEMRYKRVVSGLEFKSLQDSVYKVRSDLNARDLALYLPQSGKLYKSLVFD
APHYKSRIGIONGVOMLVDSOCYLGKDHGFTKLIMLEPQKRLPRLSILEQDE
EPPKVCORFALCSLOGRLQILSSQFOTGLIRIKKHENDNFAINEKALIRCALR
EGLKVSCEPKIOTTLRYKCFNPIPHSRSTRAFLKFGNAVILLYIORSDDIPLFL
ALAMTLKATDNLISDTSTYLIMLQGNDRYSEKLDLSGVATYDSSEPKLPMPTG
PIPAELHYTLAMDPMNVFPGYVGYLVDAEGDDLYGSQPYTYTAIIQVEERDAD
NTPFLGKIYQIDIGYSEKIVSLLDYKRSSESSONSDNAPPTSTPERLTPGLR
SIPLEFSKESHSPSTKHSRKLKYNALPILKEVYSVQAMKLEPSEKRTIIR
LYLKMHPKNNPENHIANEFKHLONEIRKLEKQAFLODNADRASRPRTSASRNS
DKYSFORPTYSNQRATSHKNSRQOOSKCKPCPSAGQYTSQEPFPPPTKSYNVEA
RMLKQARANEFAARNDLHKNANEMWCFCTYSTLALTAADYAVRAGSDKRVKPTAL
AQKIEEYSOOLGINDVHTLAVYGVDSLTRYPLDLPPOIIPNDRFSEVAMRWEC
TACIIKEINFTQOKV"

```

```

BASE COUNT 3599 a 2281 c 2387 g 3226 t
ORIGIN
Query Match 92.0%; Score 18.4; DB 10; Length 11493;
Best Local Similarity 95.0%; Pred. No. 22;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy      1  gtaaatggccacttgact 20
        |||
        6397 GTGAATGGCCACTTGTCTCT 6416

RESULT  7
G37224 434 bp DNA linear STS 31-MAR-1998
LOCUS SHGC-57260 Human Homo sapiens STS genomic, sequence tagged site.
DEFINITION G37224
ACCESSION G37224
VERSION G37224.1 GI:2996875
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 434)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLES Myers, R.M.
JOURNAL Human STS (1997)
COMMENT Unpublished
Contact: Richard M. Myers
Stanford Human Genome Center (SHGC)
Stanford University School of Medicine
Department of Genetics, M-344, Stanford, CA 94305, USA
Tel: 4157259687
Fax: 4157259689
Email: myers@shgc.stanford.edu

```

Primer A: TTTTTCAGGCGCTGTGTC
Primer B: CATGTTCTCCAGATGCCA
STS size: 99
PCR Profile:

Initial incubation: 95 degrees C for 10 minutes

Denaturation: 94 degrees C for 30 seconds
Annealing: 60 degrees C for 30 seconds
Polymerization: 72 degrees C for 23 seconds
PCR Cycles: 30
Thermal Cycler: Perkin Elmer 9700

Protocol:
Template: 25 ng
Primer: each 1 uM
dNTPs: each 200 uM
AMPLItaq Gold Polymerase: 0.07 units/uL
Total Vol: 5 uL

Buffer:
MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3

Prepared with primer pairs derived from W60474 -- Unigene.

FEATURES
source location/Qualifiers

1..434
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="10"
/clone_1b="Human"

STS
primer_bind 3..22
primer_bind complement(82..101)
BASE COUNT 98 a 93 c 97 g 146 t
ORIGIN

Query Match 87.0%; Score 17.4; DB 11; Length 434;
Best Local Similarity 94.7%; Pred. No. 69;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 tgaatggccacttgcact 20
||||| |||||||||
Db 301 TGAATGCCACTTGCACCT 319

RESULT 8
AF113358 766 bp DNA linear INV 27-APR-2000
LOCUS
DEFINITION Ips latidens haplotype 1 cytochrome oxidase I gene, partial cds;
AF113358
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mitochondrion Ips latidens
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
Cucujiformia; Phytophaga; Scolytidae; Ips.
1 (bases 1 to 766)
Cognato, A.I. and Sperling, F.A.
Phylogeny of Ips Degeer species (Coleoptera: scolytidae) inferred
from mitochondrial cytochrome oxidase I DNA sequence
MOL. Phylogenet. Evol. 14 (3), 445-460 (2000)

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
Location/Qualifiers

source

1..766
/organism="Ips latidens"
/organism="mitochondrion"
/db_xref="taxon:102838"
/haplotype="1"
<1..>766
/note="similar to Drosophila yakuba cytochrome oxidase I"
/codon_start=2
/transl_table=5
/product="cytochrome oxidase I"
/protein_id="AAF65880.1"
/translation="PGEGLSIHIGSGSKKEAFVGLMIVANTATGLGFFVVAHHM
FTIGMDVDTRAYTSATMTIATPTGKITFSWMTFTGTQISFSSISLGFLEFTM
GGTGTILANSSIDITLHDPTVYVAHFHVLSSMGATFAIAGIYQWPELFTGLTKK
YKTHPLTMEVGNLTFFPDHFGLSGMPRRSDYDAYLMMNIISSISMSLSIVF
YFIFLMEFSQAQRKSIALNLSNLSLEWLYLPSPDH"

CDS

BASE COUNT 218 a 156 c 122 g 270 t
ORIGIN

Query Match 87.0%; Score 17.4; DB 3; Length 766;
Best Local Similarity 94.7%; Pred. No. 71;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gtaatggccacttgcac 19
||||| |||||||||
Db 223 GTAATGCCACTTTCAC 241

RESULT 9
AF113359 766 bp DNA linear INV 27-APR-2000
LOCUS
DEFINITION Ips latidens haplotype 2 cytochrome oxidase I gene, partial cds;
AF113359
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mitochondrion Ips latidens
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
Cucujiformia; Phytophaga; Scolytidae; Ips.
1 (bases 1 to 766)
Cognato, A.I. and Sperling, F.A.
Phylogeny of Ips Degeer species (Coleoptera: scolytidae) inferred
from mitochondrial cytochrome oxidase I DNA sequence
MOL. Phylogenet. Evol. 14 (3), 445-460 (2000)

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1..766
/organism="Ips latidens"
/organism="mitochondrion"
/db_xref="taxon:102838"
/haplotype="2"
<1..>766
/note="similar to Drosophila yakuba cytochrome oxidase I"
/codon_start=2
/transl_table=5
/product="cytochrome oxidase I"
/protein_id="AAF65881.1"
/translation="PGEGLSIHIGSGSKKEAFVGLMIVANTATGLGFFVVAHHM
FTIGMDVDTRAYTSATMTIATPTGKITFSWMTFTGTQISFSSISLGFLEFTM
GGTGTILANSSIDITLHDPTVYVAHFHVLSSMGATFAIAGIYQWPELFTGLTKK
YKTHPLTMEVGNLTFFPDHFGLSGMPRRSDYDAYLMMNIISSISMSLSIVF
YFIFLMEFSQAQRKSIALNLSNLSLEWLYLPSPDH"

```

BASE COUNT      220 a      155 c      120 g      271 t
ORIGIN
Query Match      87.0%; Score 17.4; DB 3; Length 766;
Best Local Similarity 94.7%; Pred. No. 71;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1      tgaatgccaccttgcac 19
      |||
Db      223      GTGAATGCCACCTTTCAC 241

RESULT 10
AC069021      175309 bp      DNA      linear      HTG 30-NOV-2001
LOCUS      Homo sapiens chromosome 10 clone RP11-254K3, WORKING DRAFT
DEFINITION      AC069021
ACCESSION      AC069021
VERSION      AC069021.8      GI:17155019
KEYWORDS      HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 175309)
AUTHORS      Smith,D.R.
TITLE      Genome Therapeutics Corporation Sequencing Center: Human Genome
JOURNAL      Unpublished
AUTHORS      Smith,D.R.
TITLE      2 (bases 1 to 175309)
JOURNAL      Direct Submission
COMMENT      Submitted (17-MAY-2000) Genome Therapeutics Corporation, 100 Beaver
      Street, Waltham, MA 02453, USA
      On Nov 30, 2001 this sequence version replaced gi:14277222.
      -----
      Genome Center
      Center: Genome Therapeutics Corporation
      Center code: GTC
      Web site: http://www.genomecorp.com/
      Contact: gtc-seqcenter@genomecorp.com
      -----
      Project Information
      Center project name: hg364
      -----
      Summary Statistics
      Sequencing vector: N/A
      Chemistry: Dye-terminator Big Dye; 100% of reads
      Assembly program: Phrap; version 990315
      Consensus quality: 174470 bases at least Q40
      Consensus quality: 174909 bases at least Q30
      Consensus quality: 175102 bases at least Q20
      Insert size: 175308; sum-of-contrigs
      Quality coverage: 7.4x in Q20 bases; sum-of-contrigs
      -----
      NOTE: This is a 'working draft' sequence. It currently
      * consists of 2 contrigs. The true order of the pieces
      * is not known and their order in this sequence record is
      * arbitrary. Gaps between the contrigs are represented as
      * runs of N, but the exact sizes of the gaps are unknown.
      * This record will be updated with the finished sequence
      * as soon as it is available and the accession number will
      * be preserved.
      * 1      9697: contrig of 9697 bp in length
      * 9698      9797: gap of unknown length
      * 9798      175309: contrig of 165512 bp in length.
      Location/Qualifiers
      1. 175309
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="10"
      /clone="RP11-254K3"
      /clone_lib="RPC1-11"
      1. 9697
      /note="assembly_name:Contrig1
      misc_feature
      /note="assembly_name:Contrig1

```

```

misc_feature      clone_end:SP6"
      9798. 175309
      /note="assembly_name:Contrig2
      clone_end:77"
BASE COUNT      54898 a      36843 c      34823 g      48645 t      100 others
ORIGIN
Query Match      87.0%; Score 17.4; DB 2; Length 175309;
Best Local Similarity 94.7%; Pred. No. 91;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      2      tgaatgccaccttgcact 20
      |||
Db      14083      TGAATGCCACCTTGCACCT 14101

RESULT 11
AP000751      176713 bp      DNA      linear      PRI 22-MAR-2001
LOCUS      Homo sapiens genomic DNA, chromosome 11q, clone:RP11-679A11,
DEFINITION      complete sequence.
ACCESSION      AP000751
VERSION      AP000751.4      GI:13429919
KEYWORDS      HTG.
SOURCE      Homo sapiens DNA, clone:RP11-679A11.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (sites)
AUTHORS      Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
      Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE      Homo sapiens genomic DNA
JOURNAL      Published Only in Database (1999) In press
AUTHORS      Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
      Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE      Direct Submission
JOURNAL      Submitted (25-NOV-1999) Masahira Hattori, The Institute of Physical
      and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
      1-7-22 Suenho-cho, Tsukuba, Ibaraki, Japan, 305-0857, Japan
      (E-mail:hattori@psc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
      Tel:81-45-503-9111, Fax:81-45-503-9170)
      On Mar 21, 2001 this sequence version replaced gi:9757498.
      -----
      Location/Qualifiers
      1. 176713
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="11"
      /map="11q"
      /clone="RP11-679A11"
BASE COUNT      58670 a      33938 c      32613 g      51492 t
ORIGIN
Query Match      87.0%; Score 17.4; DB 9; Length 176713;
Best Local Similarity 94.7%; Pred. No. 91;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      2      tgaatgccaccttgcact 20
      |||
Db      68285      TGAATGCCACCTTGCACCT 68303

RESULT 12
AP001265      192913 bp      DNA      linear      HTG 30-MAY-2000
LOCUS      Homo sapiens chromosome 11 clone RP11-741L23 map 11q24, WORKING
DEFINITION      DRAFT SEQUENCE, 34 unordered pieces.
ACCESSION      AP001265
VERSION      AP001265.2      GI:8117657
KEYWORDS      HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:RP11-741L23.

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 192913)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totokl,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens 192,913 genomic DNA of 11q24
JOURNAL Published Only in DataBase (2000) In press
AUTHORS 2 (bases 1 to 192913)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totokl,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (23-FEB-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsc.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
COMMENT On May 31, 2000 this sequence version replaced gi:7106133.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: Humdraft11
Center clone name: RP11-741L23
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 170147 bases at least Q40
Consensus quality: 181282 bases at least Q30
Consensus quality: 186868 bases at least Q20
Insert size: 189613; sum-of-ctrls
Quality coverage: 4.30x in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently consists of
34 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs 'N', but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1 15621 contig of 15621 bp in length
15722 30722 contig of 15001 bp in length
30823 45308 contig of 14486 bp in length
45409 56803 contig of 11395 bp in length
56904 64499 contig of 7596 bp in length
64600 74039 contig of 9440 bp in length
74140 83223 contig of 9090 bp in length
83330 92098 contig of 8769 bp in length
92199 99531 contig of 7333 bp in length
99632 108069 contig of 8438 bp in length
108170 113891 contig of 5721 bp in length
113991 120625 contig of 6635 bp in length
120726 126591 contig of 5866 bp in length
126692 133471 contig of 6780 bp in length
133572 148668 contig of 6365 bp in length
148668 152872 contig of 6365 bp in length
152973 156332 contig of 3360 bp in length
156433 160140 contig of 3708 bp in length
160241 163773 contig of 3533 bp in length
163874 168073 contig of 4200 bp in length
168174 172134 contig of 3961 bp in length
172235 174914 contig of 2680 bp in length
174915 175014 contig of 2389 bp in length
175015 177403 contig of 2389 bp in length
177404 179082 contig of 2398 bp in length
179083 181480 contig of 2398 bp in length
181481 182602 contig of 2165 bp in length
182603 184867 contig of 2165 bp in length
184868 184968 contig of 2165 bp in length
184969 186523 contig of 1556 bp in length

184968 186523 contig of 1556 bp in length
186524 187814 contig of 1191 bp in length
187915 190302 contig of 2388 bp in length
190403 191751 contig of 1349 bp in length
191852 192913 contig of 1062 bp in length
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 34 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of 'N', but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 15621: contig of 15621 bp in length
* 15622 15721: gap of 100 bp
* 15722 30722: contig of 15001 bp in length
* 30723 30822: gap of 100 bp
* 30823 45308: contig of 14486 bp in length
* 45309 45408: gap of 100 bp
* 45409 56803: contig of 11395 bp in length
* 56804 56903: gap of 100 bp
* 56904 64499: contig of 7596 bp in length
* 64500 64599: gap of 100 bp
* 64600 74039: contig of 9440 bp in length
* 74040 74139: gap of 100 bp
* 74140 83223: contig of 9090 bp in length
* 83230 83329: gap of 100 bp
* 83330 92098: contig of 8769 bp in length
* 92099 92198: gap of 100 bp
* 92199 99531: contig of 7333 bp in length
* 99532 99631: gap of 100 bp
* 99632 108069: contig of 8438 bp in length
* 108070 108169: gap of 100 bp
* 108170 113890: contig of 5721 bp in length
* 113891 113990: gap of 100 bp
* 113991 120625: contig of 6635 bp in length
* 120626 120725: gap of 100 bp
* 120726 126591: contig of 5866 bp in length
* 126592 133471: contig of 6780 bp in length
* 133472 133571: gap of 100 bp
* 133572 139936: contig of 6365 bp in length
* 139937 140036: gap of 100 bp
* 140037 145226: contig of 5190 bp in length
* 145227 145326: gap of 100 bp
* 145327 148567: contig of 3241 bp in length
* 148568 148667: gap of 100 bp
* 148668 152872: contig of 4205 bp in length
* 152873 152972: gap of 100 bp
* 152973 156332: contig of 3360 bp in length
* 156333 156432: gap of 100 bp
* 156433 160140: contig of 3708 bp in length
* 160141 160240: gap of 100 bp
* 160241 163773: contig of 3533 bp in length
* 163774 163873: gap of 100 bp
* 163874 168073: contig of 4200 bp in length
* 168074 168173: gap of 100 bp
* 168174 172134: contig of 3961 bp in length
* 172135 172234: gap of 100 bp
* 172235 174914: contig of 2680 bp in length
* 174915 175014: gap of 100 bp
* 175015 177403: contig of 2389 bp in length
* 177404 179082: contig of 2398 bp in length
* 179083 181480: contig of 2398 bp in length
* 181481 181580: gap of 100 bp
* 181581 182602: contig of 1022 bp in length
* 182603 182702: gap of 100 bp
* 182703 184867: contig of 2165 bp in length
* 184868 184967: gap of 100 bp
* 184968 186523: contig of 1556 bp in length

FEATURES

Source

1.192913 Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="11"

/map="11q24"

/clone="RP11-741L23"

1.15621

misc_feature

/note="assembly-fragment"

15722..30722

misc_feature

/note="assembly-fragment"

30823..45308

misc_feature

/note="assembly-fragment"

45409..56803

misc_feature

/note="assembly-fragment"

56904..64499

misc_feature

/note="assembly-fragment"

64600..74039

misc_feature

/note="assembly-fragment"

74140..83229

misc_feature

/note="assembly-fragment"

83330..92098

misc_feature

/note="assembly-fragment"

92199..99531

misc_feature

/note="assembly-fragment"

99632..108069

misc_feature

/note="assembly-fragment"

108170..113890

misc_feature

/note="assembly-fragment"

113991..120625

misc_feature

/note="assembly-fragment"

120726..126591

misc_feature

/note="assembly-fragment clone_end:T7 vector_side:left"

126592..133471

misc_feature

/note="assembly-fragment"

133572..133936

misc_feature

/note="assembly-fragment"

140037..145226

misc_feature

/note="assembly-fragment"

Query Match 87.0%; Score 17.4; DB 2; Length 192913;

Best Local Similarity 94.7%; Pred. No. 92;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 tgaatggccacttgcact 20

|||||

Db 68804 TGAATGGCCACTTGCATT 68786

RESULT 13

AC073587/c

LOCUS AC073587 216521 bp DNA linear PRI 28-JUL-2001

DEFINITION Homo sapiens chromosome 10 clone RP11-572P18, complete sequence.

ACCESSION AC073587

VERSION AC073587.5 GI:15027737

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 216521)

REFERENCE Smith, D.R.

TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome

SEQUENCE DATA

JOURNAL . Unpublished

REFERENCE 2 (Bases 1 to 216521)

AUTHORS Smith, D.R.

TITLE Direct Submission

JOURNAL Submitted (25-JUN-2000) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA

REFERENCE 3 (Bases 1 to 216521)

AUTHORS Smith, D.R.

TITLE Direct Submission

JOURNAL Submitted (28-JUL-2001) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA

COMMENT On Jul 28, 2001 this sequence version replaced gi:14787184.

FEATURES

Source

1.216521 Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="10"

/clone="RP11-572P18"

/clone_lib="RPCT-11"

BASE COUNT 61858 a 44519 c 46794 g 63350 t

ORIGIN

Query Match 87.0%; Score 17.4; DB 9; Length 216521;

Best Local Similarity 94.7%; Pred. No. 92;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 tgaatggccacttgcact 20

|||||

Db 54172 TGAATGGCCACTTGCATT 54154

RESULT 14

AC019279/c

LOCUS AC019279 241392 bp DNA linear HTG 23-SEP-2000

DEFINITION Homo sapiens clone RP11-14E3, *** SEQUENCING IN PROGRESS ***; 74 unordered pieces.

ACCESSION AC019279

VERSION AC019279.5 GI:10280840

KEYWORDS HTG: HTGS_PHASE1.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 241392)

REFERENCE Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Beckert, R., Bida, F., Boguslavskiy, L., Boukhalter, B., Brown, A., Burkett, G., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearlano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galsgan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Lehocck, J., Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J., Menes, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Olivar, T. M., Peterson, K., Pierre, N., Pisanl, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A., and Zody, M.

TITLE Direct Submission

JOURNAL Submitted (31-DEC-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Sep 23, 2000 this sequence version replaced gi:7630670. All repeats were identified using RepeatMasker: <http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L3393
Center Clone name: 14_E_3

* NOTE: This is a 'working draft' sequence. It currently
* consists of 74 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
* 4479 4578: contig of 4478 bp in length
* 4479 4578: gap of 100 bp
* 4579 6006: contig of 1428 bp in length
* 6007 6106: gap of 100 bp
* 6107 7245: contig of 1139 bp in length
* 7246 7345: gap of 100 bp
* 7346 8403: contig of 1058 bp in length
* 8404 8503: gap of 100 bp
* 8504 9605: contig of 1102 bp in length
* 9606 9705: gap of 100 bp
* 9706 10965: contig of 1260 bp in length
* 10966 11065: gap of 100 bp
* 11066 12108: contig of 1043 bp in length
* 12109 12208: gap of 100 bp
* 12209 13446: contig of 1238 bp in length
* 13447 13546: gap of 100 bp
* 13547 14879: contig of 1333 bp in length
* 14880 14979: gap of 100 bp
* 14980 16230: contig of 1251 bp in length
* 16231 16330: gap of 100 bp
* 16331 17535: contig of 1205 bp in length
* 17536 17635: gap of 100 bp
* 17636 18688: contig of 1054 bp in length
* 18689 18789: gap of 100 bp
* 18790 20178: contig of 1389 bp in length
* 20179 20278: gap of 100 bp
* 20279 21831: contig of 1553 bp in length
* 21832 21931: gap of 100 bp
* 21932 23140: contig of 1209 bp in length
* 23141 23240: gap of 100 bp
* 23241 24600: contig of 1360 bp in length
* 24601 24700: gap of 100 bp
* 24701 26439: contig of 1739 bp in length
* 26440 26539: gap of 100 bp
* 26540 27663: contig of 1124 bp in length
* 27664 27763: gap of 100 bp
* 27764 29674: contig of 1911 bp in length
* 29675 29774: gap of 100 bp
* 29775 31060: contig of 1286 bp in length
* 31061 31160: gap of 100 bp
* 31161 32996: contig of 1836 bp in length
* 32997 33096: gap of 100 bp
* 33097 34353: contig of 1257 bp in length
* 34354 34453: gap of 100 bp
* 34454 36377: contig of 1924 bp in length
* 36378 36477: gap of 100 bp
* 36478 38540: contig of 2063 bp in length
* 38541 38640: gap of 100 bp
* 38641 40445: contig of 1805 bp in length
* 40446 40545: gap of 100 bp
* 40546 51430: contig of 10885 bp in length
* 51431 51530: gap of 100 bp
* 51531 53722: contig of 2192 bp in length
* 53723 53822: gap of 100 bp
* 53823 55996: contig of 2174 bp in length
* 55997 56096: gap of 100 bp
* 56097 57762: contig of 1666 bp in length

* 57763 57862: gap of 100 bp
* 57863 59701: contig of 1839 bp in length
* 59702 59801: gap of 100 bp
* 59802 61877: contig of 2076 bp in length
* 61878 61977: gap of 100 bp
* 61978 63529: contig of 1552 bp in length
* 63530 63629: gap of 100 bp
* 63630 65374: contig of 1745 bp in length
* 65375 65474: gap of 100 bp
* 65475 67237: contig of 1763 bp in length
* 67238 67337: gap of 100 bp
* 67338 68940: contig of 1603 bp in length
* 68941 69040: gap of 100 bp
* 69041 71146: contig of 2106 bp in length
* 71147 71246: gap of 100 bp
* 71247 73303: contig of 2057 bp in length
* 73304 73403: gap of 100 bp
* 73404 75790: contig of 2387 bp in length
* 75791 75890: gap of 100 bp
* 75891 78184: contig of 2294 bp in length
* 78185 78284: gap of 100 bp
* 78285 81055: contig of 2771 bp in length
* 81056 81155: gap of 100 bp
* 81156 83821: contig of 2666 bp in length
* 83822 83921: gap of 100 bp
* 83922 85940: contig of 2019 bp in length
* 85941 86040: gap of 100 bp
* 86041 88683: contig of 2643 bp in length
* 88684 88783: gap of 100 bp
* 88784 91205: contig of 2422 bp in length
* 91206 91305: gap of 100 bp
* 91306 94475: contig of 3170 bp in length
* 94476 94575: gap of 100 bp
* 94576 96917: contig of 2242 bp in length
* 96918 97017: gap of 100 bp
* 97018 99382: contig of 2365 bp in length
* 99383 99482: gap of 100 bp
* 99483 102002: contig of 2520 bp in length
* 102003 102102: gap of 100 bp
* 102103 105794: contig of 3692 bp in length
* 105795 105894: gap of 100 bp
* 105895 108858: contig of 2964 bp in length
* 108859 108958: gap of 100 bp
* 108959 113547: contig of 4589 bp in length
* 113548 113647: gap of 100 bp
* 113648 119743: contig of 6096 bp in length
* 119744 119843: gap of 100 bp
* 119844 122714: contig of 2871 bp in length
* 122715 122814: gap of 100 bp
* 122815 127700: contig of 4886 bp in length
* 127701 127800: gap of 100 bp
* 127801 131089: contig of 3289 bp in length
* 131090 131189: gap of 100 bp
* 131190 135083: contig of 3894 bp in length
* 135084 135183: gap of 100 bp
* 135184 139232: contig of 4049 bp in length
* 139233 139332: gap of 100 bp
* 139333 143122: contig of 3790 bp in length
* 143123 143222: gap of 100 bp
* 143223 148113: contig of 4891 bp in length
* 148114 148213: gap of 100 bp
* 148214 153325: contig of 5112 bp in length
* 153326 153425: gap of 100 bp
* 153426 158584: contig of 5159 bp in length
* 158585 158684: gap of 100 bp
* 158685 163480: contig of 4796 bp in length
* 163481 163580: gap of 100 bp
* 163581 168903: contig of 5323 bp in length
* 168904 169003: gap of 100 bp
* 169004 175183: contig of 6180 bp in length
* 175184 175283: gap of 100 bp
* 175284 180173: contig of 4890 bp in length
* 180174 180273: gap of 100 bp

* 180274 184306: contig of 4033 bp in length
* 184307 184406: gap of 100 bp
* 184407 189795: contig of 5389 bp in length
* 189796 189895: gap of 100 bp
* 189896 198138: contig of 8243 bp in length
* 198139 198238: gap of 100 bp
* 198239 206826: contig of 8588 bp in length
* 206827 206926: gap of 100 bp
* 206927 214548: contig of 7622 bp in length
* 214549 214648: gap of 100 bp
* 214649 221820: contig of 7172 bp in length
* 221821 221920: gap of 100 bp
* 221921 230400: contig of 8480 bp in length
* 230401 230500: gap of 100 bp

Query Match 87.0%; Score 17.4; DB 2; Length 241392;
Best Local Similarity 94.7%; Pred. No. 93;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 tgaatgacacttgcact 20
|||||
Db 79271 TGAATGCCACTTGCATT 79253

RESULT 15

POU95073

LOCUS 993 bp DNA linear INV 02-FEB-2000
DEFINITION Pisaster ochraceus cytochrome oxidase I (COI) gene, partial cds;
mitochondrial gene for mitochondrial product.

ACCESSION 995073

VERSION 095073.2 GI:6855152

KEYWORDS

SOURCE ochreous starfish.
Mitochondrion Pisaster ochraceus

ORGANISM Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Asterozoa;
Asteroidea; Forcipulatacea; Forcipulata; Asteriidae; Pisaster.

REFERENCE

AUTHORS

TITLE

1 (bases 1 to 993)
Hrincevich, A.W., Rocha-Olivares, A. and Foltz, D.W.
Phylogenetic analysis of molecular lineages in a speciose subgenus
of sea stars (Leptasterias subgenus Hexasterias)

JOURNAL

REFERENCE

AUTHORS

TITLE

2 (bases 1 to 993)
Hrincevich, A.W. and Foltz, D.W.
Direct Submission

JOURNAL

Submitted (21-MAR-1997) Zoology and Physiology, Louisiana State
University, South Campus Drive, Baton Rouge, LA 70803-1725, USA

3 (bases 1 to 993)
Hrincevich, A.W., Rocha-Olivares, A. and Foltz, D.W.
Direct Submission

Submitted (02-FEB-2000) Zoology and Physiology, Louisiana State
University, South Campus Drive, Baton Rouge, LA 70803-1725, USA

Sequence update by submitter
On Feb 2, 2000 this sequence version replaced gi:2105417.

Location/Qualifiers
1. 993

/organism="Pisaster ochraceus"
/organism="mitochondrion"
/specimen_voucher="#I-302, Louisiana State University"
/db_xref="taxon:7612"
<1. .>993
/gene="COI"
<1. .>993
/gene="COI"
/codon_start=1
/transl_table=9
/product="cytochrome oxidase I"
/protein_id="AA58051.2"
/db_xref="GI:6855153"
/translation="GNWLIPLMIGAPDMAFPMNNMFWLIPSPILLASAGVSGA
GTGWTIPPLISGSLAHAGSVLAIFSLHACASSLASINFTITINRTPGMSFDR
LPLFWMSVFTVFLDLSLPLVAGATMLTDRNINVTTFEDPAGGDPILFQHLWEF
GHPNVYILLPFGMISHVIAHAGNEPFGYLGMYAIIISIGLFLVAHMFVVG
MDVNTRAYFTATMTIAVPTGIKVFNMATLOGNNLRMDTPLLALGFEVLEFTIGILT

REMARK
COMMENT
FEATURES
source

gene
CDS

BASE COUNT 255 a 225 c 184 g 329 t
ORIGIN GVVLANSSIDIILHETYYVVAHFHYVLSMGAVFAIFAGFTHMFPLESGVSLHPLW"

Query Match 85.0%; Score 17; DB 3; Length 993;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 tgaatgacacttgcact 18
|||||
Db 733 TGAATGCCACTTGCATTGCA 749

Search completed: May 22, 2002, 06:59:46
Job time: 9171 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 08:46:53 ; Search time 3328.52 Seconds
(without alignments)
3728.214 Million cell updates/sec

Title: US-09-693-205-7_COPY_12200_12792

Perfect score: 593
Sequence: 1 acatctatgtttacaggct.....accctcacattttatgtctt 593

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vl:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*
29: em.vl:*
30: em.htg.hum:*
31: em.htg.inv:*
32: em.htg.other:*
33: em.htgo.inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description

1	593	100.0	4318	9	AB018273	AB018273 Homo sapi
2	593	100.0	12793	6	AX119931	AX119931 Sequence
3	593	100.0	12793	9	AF193556	AF193556 Homo sapi
c	593	100.0	92693	9	AL157766	AL157766 Human DNA
5	593	100.0	99819	2	AC079761	AC079761 Homo sapi
c	155	26.1	418	11	G36555	G36555 SHGC-53325
7	96	16.2	3289	9	AB056815	AB056815 Macaca fa
8	51	8.6	51	6	AX160575	AX160575 Sequence
9	51	8.6	51	6	AX160577	AX160577 Sequence
10	34	5.7	50	6	AX160578	AX160578 Sequence
11	25	4.2	51	6	AX160576	AX160576 Sequence
12	24	4.0	50	6	AX160574	AX160574 Sequence
13	24	4.0	51	6	AX160573	AX160573 Sequence
14	23	3.9	149128	2	AL670276	AL670276 Mus muscu
15	23	3.9	336028	2	AC092096	AC092096 Mus muscu
c	16	3.7	349980	6	AX344559	AX344559 Sequence
17	21	3.5	1557	4	AF018634	AF018634 Canis fam
c	18	3.5	4604	6	AX278036	AX278036 Sequence
c	19	3.5	4604	6	AX333807	AX333807 Sequence
c	20	3.5	4604	6	AX344874	AX344874 Sequence
c	21	3.5	35920	3	CEC4484	CEC44874 Sequence
21	21	3.5	38418	10	M0SPRNP4	Z79598 Caenorhabdl
c	22	3.5	39760	10	U29187	U29186 Mus musculu
c	23	3.5	70892	2	AC027533	AC027533 Mus muscu
c	24	3.5	71456	10	MMA298054	AC027533 Homo sapi
c	25	3.5	74102	9	AL356316	AL356316 Homo sapi
c	26	3.5	79946	9	AL359874	AL359874 Human DNA
c	27	3.5	110000	2	PFMAL13P2_1	Confinuaction (2 of
c	28	3.5	141794	2	AC093455	AC093455 Homo sapi
29	21	3.5	149901	2	AC022243	AC022243 Homo sapi
30	21	3.5	168957	2	AC025079	AC025079 Homo sapi
c	31	3.5	171279	2	AC036185	AC036185 Homo sapi
c	32	3.5	171271	9	AC022821	AC022821 Homo sapi
33	21	3.5	172639	30	AC009566	AC009566 Homo sapi
34	21	3.5	176333	2	AC108022	AC108022 Homo sapi
c	35	3.5	176863	2	AL358353	AL358353 Homo sapi
c	36	3.5	178100	9	AC009080	AC009080 Homo sapi
c	37	3.5	180526	9	AC027333	AC027333 Homo sapi
38	21	3.5	182725	2	AC021206	AC021206 Homo sapi
39	21	3.5	184529	9	AC009156	AC009156 Homo sapi
40	21	3.5	187047	9	AC018861	AC018861 Homo sapi
41	21	3.5	197551	2	AL358073	AL358073 Homo sapi
c	42	3.5	206187	2	AC069139	AC069139 Homo sapi
43	21	3.5	245997	2	AC093360	AC093360 Mus muscu
44	21	3.5	245997	2	AC093360	AC093360 Mus muscu
45	20	3.4	1947	9	AK023911	AK023911 Homo sapi

ALIGNMENTS

RESULT	1	4318 bp	mrna	linear	PRI 16-JUN-1999
LOCUS	AB018273				
DEFINITION	Homo sapiens mRNA for KIAA0730 protein, partial cds.				
ACCESSION	AB018273				
VERSION	AB018273.1	GI:3862180			
KEYWORDS					
SOURCE					
ORGANISM	Homo sapiens adult male brain cDNA to mRNA, clone_11b:pbluescriptII SK plus clone:hk03632.				
REFERENCE					
AUTHORS	1 (sites)				
TITLE	Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.				
JOURNAL	Prediction of the coding sequences of unidentified human genes. XI. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro				
MEDLINE	DNA Res. 5 (5), 277-286 (1998)				
REFERENCE	99087487				
AUTHORS	2 (bases 1 to 4318)				
TITLE	Ohara,O., Suyama,M., Nagase,T., Ishikawa,K. and Kikuno,R.				
	Direct Submission				

JOURNAL

Submitted (08-Oct-1998) Osamu Ohara, Kazusa DNA Research Institute,
Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba
292-0812, Japan (E-mail:cdhainfo@kazusa.or.jp, Tel:+81-438-52-3913,
Fax:+81-438-52-3914)

FEATURES

Location/Qualifiers

1..4318
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="h03632"
/sex="male"
/tissue_type="brain"
/clone_lib="pbluescriptII SK plus"
/dev_stage="adult"
1..3015
/gene="K1AA0730"
<1..3015
/gene="K1AA0730"
/codon_start=1
/product="K1AA0730 protein"
/protein_id="BAA34450.1"
/db_xref="GI:3882181"
/translation="VTFTMSWVEFLRNIGLAKYILSSQOOLQFRAKEISVANTEMNSKE
TLQNTVDILHHRIOBRDLISGNFLKELSLIPICPEPAEPIFRHQVEVNTL
PLIKNGAONPKFKOCVQLQMLTSCPLPEKATPLSTKEDSGDLGQEDLEVLN
MLNVLDPDLKVLINNCNINICITTLDEEMVTRAKVLKSTIEFLSAEKREPRFLQNG
VAFVVEDGWLKLKEEVLINLEYSDEKPYLYKPLLEGLTFHQLKLGTDIISTK
QYEVLSIFKNSKGLDPEMRYKVSGLFSLKNDYSVCFKHGFPETKLMLEPO
PSODRLVKSILVEFDAPHYKSRIOGNVOMLDSOCYGHGHPETKLMLEPO
KLPRLLSILIEEOLDEFPKVCORGALCSIOGRLOLLSSQFPTGLIRIKHNDN
AFLANEKRIKRCALRGLKSKCEKQCTTLRVGPNPIPRSRSETRAFLRGNV
ILYI0HSDSKDINFLALAMTLKSAIDNLSDSYLYLMLCNDIYRIGELDLSYG
KYDSEPSKLELPMGCTPIPAEIHYYTLMDPNVYRPGRYVGLVDAEGDIDYGYOP
TYYTAIIQVEVEDADNSFLKTYQIDIGYSEKIVSLDYKFSRPEESGSQDS
ASTPTSEPLETPLGRSLPPLSGRSHKTSKQSPKLVNSLPELTKFVTSVE
QAMKIPESRKKIIRRLYLKMPDKNEMHDIANVFKHONEIRLEQALDNAD
BASRSTESASRSDSKYSPORFYSNNQENTSKSKEQOQNKCKPSPACQYTSOR
FVPTPEKSVGNPVEARRMLRORANFSARNDLHKNAEMWCFRCYLSIKLALIAAD
YAVRKSQDKVPTALQKIEEYSOOLBGLNVDHVTLEAYGVDSLKTRYPDLLPPOI
PMDRPTSEVAMVMECTACIIIKLENFMQKV"

CDS

gene

BASE COUNT 1395 a 734 c 847 g 1342 t
ORIGIN

Query Match 100.0%; Score 593; DB 9; Length 4318;
Best Local Similarity 100.0%; Pred. No. 1.3e-308;
Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acatctatggttaacaggtctcctggttgaatgaagatagaacggaacacaaatggt 60
|||||
DB 3650 ACATCTTATGTTTACAGGCTTCCTGTTGATGAAAGATAGCAAGAAACTCAAAATGGT 3709
QY 61 ggcagttctatataccagctgttagtatgtttctcgaaactcctgcgaagaacacatt 120
|||||
DB 3710 GGCAGTCTTATATACAGATGTTAGTATGTTCTCGAANACTGCTGCCAAGCAACATT 3769
QY 121 taataacgttagaaccactgcttactgttctgtgtgtacataattccacaatgttata 180
|||||
DB 3770 TATTAACTGTTAGAACACTGCTTATGTTGTTGCTACATATTTCCACAATGTTATA 3829
QY 181 attatatagttgtgtgaacagatgatacttttctgtctaaagtgctgcagttaa 240
|||||
DB 3830 ATTATATAGTGTGTGTAACAGAGATGCAATCTTTGTTGTCTAAAGGCTGCACGTTAA 3889
QY 241 aaaaaaacaacacctttcttcaatataggcatgttagtgagttttttaaacttaaaac 300
|||||
DB 3890 AAAAAAACAACCTTTCTTTCAATATGGCATGTAGTGAAGTTTTTTTAACTTAAAAAC 3949
QY 301 atcaaaaatgtttaaaatcatgtgttactagtagttataaataatcgactataatc 360
|||||
DB 3950 ATCAAAAATGTTAAATCATGTTGTATCTAGTATTATATATCGGCTTATATTTTC 4009
QY 361 cccatgaatgatcagaacatgaacttaattcattgttctcgcacgtcttcttactt 420
|||||

DB 4010 CCCATGAATGATCAGACATGACATTTAATTCATGTTTGTCTGCCATGCTTCTTACTTT 4069

QY 421 aacataattcttctgagaatgtaaaaggttaagataatgaattatataagtgactcg 480
|||||

DB 4070 AACATATTTCTTTTGCAGAAATGTAAGGTAAGTAATTAATTTAGTTTATTAAGGTCCTGG 4129

QY 481 ctgtaaatgctctaaataacttataatgaatgaaggtcttaagaacatgttgaaact 540
|||||

DB 4130 CTGTAAATGATGCTTAATATATCTTATGCAATTAAGGCTTACAGAAACATGTTGAACCTT 4189

QY 541 ttttcttcttattgggaatgaagatgttttgcaactccaactttatgctt 593
|||||

DB 4190 TTTTACTTTTATTTGGCAATTAAGGAATGTTTGCACCTCCACATTTTATTTGCTT 4242

RESULT 2

AX119931 2
LOCUS AX119931 12793 bp DNA linear PAT 11-MAY-2001

DEFINITION Sequence 1 from Patent WO0129266.

ACCESSION AX119931

VERSION AX119931.1 GI:14036678

KEYWORDS

SOURCE

ORGANISM

human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 12793)
Hudson,T.J., Engert,J. and Richter,A.
Identification of areas mutations and methods of use therefor
Patent: WO 012926-A 1 26-APR-2001;

McGILL UNIVERSITY (CA) ; Hopital Sainte-Justine (CA)
Location/Qualifiers

FEATURES

source

1..12793
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 4163 a 2256 c 2487 g 3887 t
ORIGIN

Query Match 100.0%; Score 593; DB 6; Length 12793;
Best Local Similarity 100.0%; Pred. No. 1.3e-308;
Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acatctatggttaacaggtctcctggttgaatgaatagaacggaacacaaatggt 60
|||||
DB 12201 ACATCTTATGTTTACAGGCTTCCTGTTGATGAAAGATAGCAAGAAACTCAAAATGGT 12260
QY 61 ggcagttctatataccagctgttagtatgtttctcgaaactcctgcgaagaacacatt 120
|||||
DB 12261 GGCAGTCTTATATACAGATGTTAGTATGTTCTGGAACCTGCTGCCAAGCAACATT 12320
QY 121 taataactttaagaaccactgcttactgttctgtgtgtgtacataattccacaatgttata 180
|||||
DB 12321 TATTAACTGTTAGAACACTGCTTATGTTGTTGCTACATATTTCCACAATGTTATA 12380
QY 181 attatatagttgtgtgaacagatgatacttttctgttctaaagtgctgcagttaa 240
|||||
DB 12381 ATTATATAGTGTGTGTAACAGAGATGCAATCTTTGTTGTCTAAAGGCTGCACGTTAA 12440
QY 241 aaaaaaacaacacctttcttcaatataggcatgttagtgagttttttaaacttaaaac 300
|||||
DB 12441 AAAAAAACAACCTTTCTTTCAATATGGCATGTAGTGAAGTTTTTTTAACTTAAAAAC 12500
QY 301 atcaaaaatgtttaaaatcatgtgttactagtagttataaataatcgactataatc 360
|||||
DB 12501 ATCAAAAATGTTAAATCATGTTGTATCTAGTATTATATATCGGCTTATATTTTC 12560
QY 361 cccatgaatgatcagaacatgaacttaattcattgttctcgcacgtcttcttactt 420
|||||
DB 12561 CCATATGATGATCAGACATGACATTTAATTCATGTTTGTCTGCCATGCTTCTTACTTT 12620
QY 421 aacataattcttctgagaatgtaaaaggttaagataatgaattatataagtgactcg 480
|||||

|||||
Db 12441 AAAAAAAAAACCTTTCTTCAATATGCATGAGTGAGTCTTTTAACTTAAAAAC 12500
QY 301 atcaaaaatgtaaaatcctgtgtatctagttaattatcagcttatattc 360
Db 12501 ATCAAAAATGTTAAATCAATGCTTATCTAGTATTAATATCGGCTTATATTC 12560
QY 361 cccatgaatgacgaactacattatcatctgtctcgcacgtcttcttacttc 420
Db 12561 CCCATGAAATATGACGAACTGACATTATCATGTTGCTCGCAATGCTTCTTACTTT 12620
QY 421 aacatattctttgcgaagttaaaaggttaagataattagttatataagtgactg 480
Db 12621 AACATATTCTTTGCGAAGTAAAGGTAAATGATTAATTAGTTATATAGTACTCG 12680
QY 481 ctgtaaatgtgcttaaatatcttattatgcattaaaggtctacagacaagtgaaact 540
Db 12681 CTGTAATGATGCTAAATTAATCTTATGCAATTAAGGCTTACAGAACTGTGAAACTT 12740
QY 541 ttcttacttattgggaataagaatgtcttcgacccctccacttatttgcct 593
Db 12741 TTTTACTTTTATTTGGGAATAGGAATGTTGCACTCCACATTTTATTTGCTT 12793
RESULT 4
AL157766/c 92693 bp DNA linear PRI 11-Apr-2001
LOCUS Human DNA sequence from clone RP11-40020 on chromosome
DEFINITION 13q12.11-12.2, complete sequence.
ACCESSION AL157766
VERSION AL157766.9 GI:13620292
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 92693)
AUTHORS Tromans,A.
TITLE Direct Submission
JOURNAL Submitted (11-Apr-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Apr 12, 2001 this sequence version replaced gi:12709868.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TrEMBL; Wp: WormPep; Information on the WormPep
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
RP11-40020 is from the library RPc1-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: PBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-40020. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP11-760M1 is at 92594 in this sequence.

FEATURES The true right end of clone RP11-72P19 is at 100 in this sequence.
source Location/Qualifiers
1..92693
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="q12.11-12.2"
/clone="RP11-40020"
/clone_1lb="RPc1-11.1"
2390..2485
/note="MIR repeat: matches 81..192 of consensus"
2562..2673
/note="MIR repeat: matches 68..183 of consensus"
3896..4201
/note="AluY repeat: matches 3..308 of consensus"
5122..5397
/note="MER46C repeat: matches 1..286 of consensus"
18986..19294
/note="AluY repeat: matches 1..310 of consensus"
19644..19873
/note="MER46A repeat: matches 1..235 of consensus"
20613..20912
/note="AluSg1 repeat: matches 1..300 of consensus"
23342..23651
/note="AluSg1 repeat: matches 1..309 of consensus"
24769..24891
/note="L2 repeat: matches 2554..2682 of consensus"
25871..26011
/note="L2 repeat: matches 2356..2495 of consensus"
26033..26109
/note="L2 repeat: matches 2601..2688 of consensus"
26245..26344
/note="L2 repeat: matches 2154..2255 of consensus"
26938..27096
/note="MIR repeat: matches 3..175 of consensus"
27150..27653
/note="L2 repeat: matches 1063..1644 of consensus"
28522..28891
/note="THE1B repeat: matches 1..364 of consensus"
29447..29834
/note="LIME3A repeat: matches 5787..6164 of consensus"
36098..36415
/note="AluSx repeat: matches 1..308 of consensus"
37202..37414
/note="MIR repeat: matches 22..262 of consensus"
37963..38254
/note="AluSg repeat: matches 9..301 of consensus"
38703..39008
/note="AluSg repeat: matches 1..306 of consensus"
39790..40093
/note="AluSx repeat: matches 1..304 of consensus"
40126..40416
/note="AluSg repeat: matches 1..292 of consensus"
40444..40733
/note="AluSg repeat: matches 1..292 of consensus"
41322..41405
/note="Single clone region. Assembly confirmed by
restriction digest data."
41541..41788
/note="AluSg repeat: matches 1..248 of consensus"
44790..45101
/note="AluSg repeat: matches 1..313 of consensus"
45261..45312
/note="13 copies 4 mer tggc 88% conserved"
45899..46206
/note="AluY repeat: matches 1..307 of consensus"
46754..47052
/note="AluY repeat: matches 1..298 of consensus"
47067..47365
/note="AluY repeat: matches 1..299 of consensus"
47477..47873
/note="L1MA10 repeat: matches 5950..6322 of consensus"
47889..48229

repeat_region /note="AlusX repeat: matches 1. .312 of consensus"
49168. .49212
/note="Alu repeat: matches 85. .126 of consensus"
repeat_region 49620. .49693
/note="L2 repeat: matches 1685. .1757 of consensus"
repeat_region 50704. .51032
/note="AlusC repeat: matches 1. .308 of consensus"
misc_feature 52204. .53009
/note="CpG island"
evidence=not_experimental
repeat_region 53978. .54137
/note="AlusX repeat: matches 1. .160 of consensus"
repeat_region 54179. .54511
/note="LMB6 repeat: matches 5822. .6172 of consensus"
repeat_region 55685. .55949
/note="AlusX repeat: matches 1. .305 of consensus"
repeat_region 57331. .57390
/note="30 copies 2 mer ga 75% conserved"
repeat_region 57357. .57392
/note="9 copies 4 mer gaga 91% conserved"
repeat_region 57753. .57930
/note="MIR repeat: matches 82. .262 of consensus"
repeat_region 58260. .58389
/note="MIR repeat: matches 2. .153 of consensus"
repeat_region 58564. .58611
/note="24 copies 2 mer ca 93% conserved"
repeat_region 59350. .59533
/note="AlusG repeat: matches 129. .313 of consensus"
repeat_region 59922. .60223
/note="AluJo repeat: matches 85. .299 of consensus"
repeat_region 61036. .61144
/note="L2 repeat: matches 2581. .2696 of consensus"
repeat_region 62008. .62187
/note="TIGER1 repeat: matches 2238. .2418 of consensus"
repeat_region 62188. .62316
/note="AluY repeat: matches 1. .129 of consensus"
repeat_region 62330. .62363
/note="Alu repeat: matches 261. .294 of consensus"
repeat_region 62362. .62565
/note="TIGER1 repeat: matches 1586. .1787 of consensus"
repeat_region 62566. .62865
/note="AlusQ repeat: matches 1. .302 of consensus"
repeat_region 62866. .64385
/note="TIGER1 repeat: matches 46. .1586 of consensus"
repeat_region 64386. .64694
/note="AluY repeat: matches 1. .306 of consensus"
repeat_region 64695. .64713
/note="TIGER1 repeat: matches 29. .46 of consensus"
repeat_region 65068. .65395
/note="LIPB2 repeat: matches 5405. .5733 of consensus"
repeat_region 65396. .65569
/note="AluY repeat: matches 136. .309 of consensus"
repeat_region 65571. .65640
/note="LIPB2 repeat: matches 5728. .5791 of consensus"
repeat_region 65696. .65717
/note="11 copies 2 mer ta 100% conserved"
repeat_region 65725. .66096
/note="LIPB2 repeat: matches 5789. .6155 of consensus"
repeat_region 66371. .66410
/note="10 copies 4 mer tgcg 82% conserved"
repeat_region 67586. .67886
/note="AluB repeat: matches 1. .299 of consensus"
repeat_region 69748. .69930
/note="MIR repeat: matches 6. .248 of consensus"
repeat_region 70957. .71267
/note="AluY repeat: matches 1. .311 of consensus"
repeat_region 71279. .71413
/note="MER21B repeat: matches 548. .680 of consensus"
repeat_region 71411. .71737
/note="MER31A repeat: matches 47. .485 of consensus"
repeat_region 71780. .72075
/note="AlusX repeat: matches 1. .295 of consensus"
repeat_region 72145. .72256

repeat_region /note="MER31-internal repeat: matches 42. .175 of consensus"
72454. .72865
/note="MER31-internal repeat: matches 332. .739 of consensus"
repeat_region 72873. .73249
/note="MER31-internal repeat: matches 883. .1261 of consensus"
Query Match 100.0%; Score 593; DB 9; Length 92693;
Best Local Similarity 100.0%; Pred. No. 1.3e-308; Mismatches 0; Gaps 0;
Matches 593; Conservative 0; Mismatches 0; Indels 0;
OY 1 acatctatgttaccagctctcgtttgtagaagatagacaaggaacacaaatcgt 60
|||||
Db 6709 ACATCTTATGTTTACAGGCTTCTGTTGATGAAGATAGCAGCGAAATCAAAATGCT 6650
|||||
OY 61 ggcagcttcttaccagcttctgtagtattgttctggaactgttgcgaagaacatt 120
|||||
Db 6649 GGCAGTCTTATACAGTGTGTAGTATGTTCTGGAACCTGTCGCAAGACACATT 6590
|||||
OY 121 tattaactgttagaacaactgtcttatgttgtgtgtacatatctccacaatgtcata 180
|||||
Db 6589 TATTAACTGTAGAACACTTCTTATGTTGTGTACATATTTTCCACAAATGTATATA 6530
|||||
OY 181 attatatagttgttgaagaagatgcacattcttctgtctaaagtgctgcagttaa 240
|||||
Db 6529 ATTATATAGTGTGTGAACGAGATGCATCTTTGTGTAAAGTGCTGCAGTTAA 6470
|||||
OY 241 aaaaaaacaaccccttcttctcaatatagcacatgtagtggagtttcttaacttaaaac 300
|||||
Db 6469 AAAAAAACAAACCTTTCTTCTTCAATATGCGCATGTAGTGGAGTTTAACTTAAAC 6410
|||||
OY 301 atcaaaaatgtttaaataatcattgtgtatctagtagtattataatcgcgttatctc 360
|||||
Db 6409 ATCAAAAATGTTAAATCATGTGTATGTATGATGTTTAAATATCGGCTTAAATTC 6350
|||||
OY 361 cccatgaatgaacgaactgcacattatcaatgattgtgtcgcgcagctcttacttc 420
|||||
Db 6349 CCCATGAATGATCAAACTGCATTTAATTCATGTGTCTCGCAGTCTTACTT 6230
|||||
OY 421 aacatattcttcttcagaaatgtaaagaaatgataatgattatataagtgctacg 480
|||||
Db 6289 AACATATTTCTTTCGAGAATGTAAAGGTAAATGATATATGTTTAAATAGTACTCG 6230
|||||
OY 481 ctgtcaaatgtagtcaaatataacttataatgcaatgaaggtctacgaacaatgtaact 540
|||||
Db 6229 CTGTAAATGATGCTAAATATATCTTATGCAATTAAGGCTTACGAACATGTTAAACTT 6170
|||||
OY 541 ttctactttattgggaataaggaatgattgtgcacccacacatttactgct 593
|||||
Db 6169 TTTTACTTTTATGGGAATAGGAATGTTTGACCTCCACATTTTATTCCTT 6117
|||||
RESULT 5
AC079761 99819 bp DNA linear HTG 10-SEP-2000
LOCUS
DEFINITION Homo sapiens chromosome UNK clone RP11-143G17, *** SEQUENCING IN
PROGRESS ***, 44 unordered pieces.
ACCESSION AC079761
VERSION AC079761.1 GI:10047966
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 99819)
AUTHORS Waterston, R. H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 99819)
AUTHORS Waterston, R. H.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-2000) Genome Sequencing Center, Washington


```

/misc_feature /note="assembly_name:Contig57"
28879..30893
/misc_feature /note="assembly_name:Contig58"
30994..32460
/misc_feature /note="assembly_name:Contig59"
32561..33984
/misc_feature /note="assembly_name:Contig60"
34085..35285
/misc_feature /note="assembly_name:Contig61"
35386..37184
/misc_feature /note="assembly_name:Contig62"
37285..39172
/misc_feature /note="assembly_name:Contig63"
39273..40874
/misc_feature /note="assembly_name:Contig64"
40975..42893
/misc_feature /note="assembly_name:Contig65"
42994..44384
/misc_feature /note="assembly_name:Contig66"
44485..45999
/misc_feature /note="assembly_name:Contig67"
46100..48669
/misc_feature /note="assembly_name:Contig68"
48770..50798
/misc_feature /note="assembly_name:Contig69"
50899..52809
/misc_feature /note="assembly_name:Contig70"
52910..55127
/misc_feature /note="assembly_name:Contig71"
55228..58087
/misc_feature /note="assembly_name:Contig72"
58188..61004
/misc_feature /note="assembly_name:Contig73"
61105..64185
/misc_feature /note="assembly_name:Contig74"
64286..67105

```

Query Match 100.0%; Score 593; DB 2; Length 99819;

Best Local Similarity 100.0%; Pred. No.1.3e-308;

Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 acatctatgtttacaagcttcctgcttgcgttgaagaagcaagcaaacctcaaatggt 60
|||||
Db 61122 ACATCTATGTTTACAGGCTTCCTGTTGATGAGATAGCAGCAAACTCAAAATGCT 61181

Qy 61 ggcagttcttataccagttgttagtattgtttctggaactgcttgccaagacaact 120
|||||
Db 61182 GGCAGTCTTATACAGTGTGTGATGTGTTCTGGAACCTGCTGCAGAACACATTT 61241

Qy 121 tattaactgttagaaccttgctttagtctgtgtgtacataatttccacaatgtcata 180
|||||
Db 61242 TATTAACTGTAGAACCTGCTTATGTGTGTGTGATATTTCCACAAATGTTATA 61301

Qy 181 attataatgttggttgacaagaatgcaatctttgttctaaagtgctgcaatgaa 240
|||||
Db 61302 ATTATATATGTTGTTGAACAGATGCATCTTTTGTGCTAAAGTGTGCAATTTAA 61361

Qy 241 aaaaaaaacaacctttcttccaatagtcagtagtgcagtttctttaaactttaa 300
|||||
Db 61362 AAAAAAACAACCTTTCTTCAATATGCGATGTAGTGAGTTTAACTTTAAATAAAC 61421

Qy 301 atcaaaaatgttaaaatcaatgtgtatctagtagttataatcatcgactataatc 360
|||||
Db 61422 ATCAAAAATGTTAAATCAATGTGTATCTAGTATCTTAATTAATCGGCTTATATTC 61481

Qy 361 cccatgaatgacgaagctacattatcatgcttgcgtcgcgaagctcttcttacttt 420
|||||
Db 61482 CCCATGATGATCAGAACTGACATTTATTCATCTTTGTCTCGCAGTCTTCTTACTTT 61541

Qy 421 aacataatcttctgcagaatgttaaaagtataatagttataatagttactgag 480
|||||
Db 61542 AACATATTTCTTTTGAGAAATGTAAAGGTAAATAGTATTAATTAAGTCTACTGG 61601

```

```

Qy 481 ctgtaaatgctgcaaatatacttataatgcaatgaagggtctacagaacatgtgaact 540
|||||
Db 61602 CTGTAATGATGCTAAATATTAATCTTTATGCAATTAAGGCTTACGAAACATTTAAACTT 61661

Qy 541 ttttactttatgtggaataaagaatgttgcacccctcacattttatgctt 593
|||||
Db 61662 TTTTACTTTTATTTGGCAATTAAGCAATGTGTGACACCTCCACATTTTATTTGCTT 61714

```

```

RESULT 6
G36555/c 418 bp mRNA linear STS 31-DEC-1997
LOCUS SHGC-53325 Human Homo sapiens STS cDNA, sequence tagged site.
DEFINITION ACCSSION G36555
VERSION G36555.1 GI:2734222
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 418)
AUTHORS Myers, R.M.
TITLE Human STS (1997)
JOURNAL Unpublished
COMMENT
Contact: Richard M. Myers
Stanford Human Genome Center (SHGC)
Stanford University School of Medicine
Department of Genetics, M-344, Stanford, CA 94305, USA
Tel: 4157259687
Fax: 4157259689
Email: myers@shgc.stanford.edu
Primer A: GTGAGGTGCAACATTCCT
Primer B: ACATTTAATCATGTTGTGCCG
STS size: 202
PCR Profile:
Initial incubation: 95 degrees C for 10 minutes
Denaturation: 94 degrees C for 30 seconds
Annealing: 60 degrees C for 30 seconds
Polymerization: 72 degrees C for 23 seconds
PCR Cycles: 30
Thermal Cycler: Perkin Elmer 9600
Protocol:
Template: 25 ng
Primer: each 1 uM
dNTPs: each 200 uM
Ampliflag Gold Polymerase: 0.07 units/ul
Total Vol: 5 ul
Buffer:
MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3

```

Prepared with primer pairs derived from M46342 -- Unigene.
Location/Qualifiers
1..418
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="13"
/clone_lib="Human"
STS
89..290
primer_bind complement(268..290)
BASE COUNT 167 a 65 c 55 g 130 t 1 others
ORIGIN

Query Match 26.1%; Score 155; DB 11; Length 418;
Best Local Similarity 99.2%; Pred. No.2.7e-72;
Matches 255; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	253	ctttctcttccaatgagcatgtagtggaggttttttaactttaaaaacatccaaaattgt	312
Db	418	CTTTCTTTCAATAGGCGATGAGCGGAGTTTTTTTAACTTTAAAAACATCAAAATTTCT	359
QY	313	taaaatcatctgltatcatagtagttaaataatlcgsgcttaataattcccaatgaatgat	372
Db	358	TAAATCAATGTGGTATCTAGTACTTTPAATTAATTAACGGCTATATTTCCCATGAAATGAT	299
QY	373	cagaactcaactttaatcatggttgtctgcgcagctcttcttaactttaaacattctt	432
Db	298	CAGAACTGACATTTATTCATGTTGTGTCGCCGCAAGCTCTTTACTTTAACTAATTTCTT	239
QY	433	ttgcagaaatgtaaaagyaatgataataatagttatataataagtgtaactggtgtaaatgatg	492
Db	238	TTGCGAATGTAAAAAGGTAAATGATTAATTAATTAATTAAGTACTGCGTGAATATGATG	179
QY	493	ctaaataacttaacttgc	509
Db	178	CTAAATATACCTTTATGC	162

RESULT	7	
AB056815		
LOCUS	3289 bp	mRNA
DEFINITION	Macaca fascicularis brain cDNA clone:Qf1A-15307, full insert sequence.	linear PRI 14-MAR-2001
ACCESSION	AB056815	
VERSION	AB056815.1	GI:13365931
KEYWORDS	fts (full insert sequence); oligo capping.	
SOURCE	Macaca fascicularis adult male frontal lobe	
ORGANISM	Macaca fascicularis	
	clone_11b:macaque brain cDNA library Qf1A clone:Qf1A-15307.	

REFERENCE	AUTHORS	TITLE
1 (sites)	Osada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hirai, M., Terao, K., Suzuki, Y., Sugano, S. and Hashimoto, K.	Isolation of full-length cDNA clones from macaque brain cDNA libraries

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3289)
AUTHORS Hashimoto, K., Osada, N., Hida, M., Kusuda, J. and Sugano, S.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2001) Katsuyuki Hashimoto, National Institute of

COMMENT

Vector: pME18S-FL3 (Acc. No. AB009864)
R. Site1: DraIII (CAGCTGTG)
R. Site2: DraIII (CAGCATGG)
Description: 1st strand cDNA was primed with an oligo(dT) primer using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of pME18S-FL3. XhoI sites just outside the DraII sites can be used to isolate the cDNA insert. Libraries were constructed by Sugano et al. (University of Tokyo, Institute of Medical Science). Custom primer used for sequencing
(5' end primer (CTTCTGCTCTTAAGAGTGG) ;
3' end primer (CGACCTGACGCTGAGGACCA)).

FEATURES

```

/organism="Macaca fascicularis"
/db_xref="taxon:9511"
/clone="Qf1a-15307"
/sex="male"
/tissue_type="frontal lobe left"
/clone_lib="macaque brain cDNA library Qf1a"

```

```

/dev_stage="adult"
/notes="Host: TOP10; Vector: pME185-FLJ (Acc.No. AB009864)
R. Site1: DraIII (CACGCTG); R. Site2: DraIII
(CACCATGTG)"
35..1951
CDS

```

```

/codon_start=1
/product="hypothetical protein"
/protein_id="BAB39340.1"
/db_xref="GI:13365932"
/translation="MLVDSQCYLGKDHGFHTKLMLPQKLRPRLSSILEEQIDEE
TPKVGCFALCSLGRLQILSSSEQFIIGLIRIMKHEDNFMLEEKARICRLKRL
GLKVCSEKLOTLLRVKGFNPISPHRSSTPAIRKFGNAVILLYOHSQSDINLLR
LAMLKSKTDNLDISTSLIAMLGNDIYRIGEKLDLSAGVYIDSESPKLEIPMGCTI
IPAEIHYLLMDPMNVFPGEYVGLVDADEGDIYGSQPTVPAIIVQEEDEADAD
SSFGKIIYQIDIGYSEYKIVSLDLYKSDREDDOSSQSTPTTAYSTEFELAPLR
IPLPESGESHKTSKHQSPKKLNVNSLPETLKEVTSVVEADAMKLPSEERKIIRLI
LKMPDKPNENHDIANEYFKRLQNEINLEKQAFLDQNAADASRTFETSASROSDY
YSFQRYTSMNQEATSHKSERQOQKKECPQSAQTVSQRFPVPTFESVGNAPPEAR
WLRARANFSAARDNLKHNANEWCFCYKLTSLALVADYAVRKSQDKVPTPLAAK
KIEESQOLEGTLVDVHTLEAYGVDSLKTRYPDLLPQIIPNDRTSEVARVMECTY
CIILKENFIQKY"

```

BASE COUNT	1062 a	552 c	653 g	1022 t
ORIGIN				
Query Match	16.2%	Score 96;	DB 9;	Length 3289;
Best Local Similarity	99.3%;	Pred. No.	1.8e-40;	
Matches 146;	Conservative	0;	Mismatches 1;	Indels 0;
				Gaps 0;

Oy	92	ttctcgtgaacatcctctgcgaagacacacattatctaactgtgtgaacacctgtctctatgtct	151
Db	2684	ttctcgtgaacatcctctgcgaagacacacattatctaactgtgtgaacacctgtctctatgtct	2743
Oy	152	gtgtgtacatattttccacacaaatgttatataatgtagtgtgtgtgaacagatgtgaat	211
Db	2744	gtgtgtatattattttccacacaaatgttatataatgtagtgtgtgtgaacagatgtgaat	2803
Oy	212	ctttctgtgtctaaagtgctgcagct	238
Db	2804	ctttgtgtgtctttaaagtgctgtccagct	2830

RESULT	8
AX160575	
LOCUS	AX160575
DEFINITION	Sequence 3903 from Patent WO0440521.
ACCESSION	AX160575
VERSION	AX160575.1 GI:14541906
KEYWORDS	
SOURCE	human.

ORGANISM	Homo sapiens
Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 51)

AUTHORS Shimkets, R.A. and Leach, M
TITLE Nucleic acids containing single nucleotide polymorphisms and methods of use thereof
JOURNAL Patent: WO 0140521-A 3903 07-JUN-2001;

FEATURES	Curagen Corporation (US)
source	Location/Qualifiers
	1. .51

```
misc_feature      /organism="Homo sapiens"  
                  /db_xref="taxon:9606"  
note="1 of 2 allelic variants (3904 is other entry)"  
Accession number  cg43924289  
BASE COUNT      15 a      8 c      8 g      20 t  
ORIGIN
```

```

Query Match      8.6%:  Score 51:  DB 6:  Length 51:
Best Local Similarity 100.0%:  Pred. No. 3.5e-16:
Matches 51:  Conservative 0:  Mismatches 0:  Indels 0:  Gaps 0:

```

Oy 109 caagacacatttactttagaaccttcttattgtgtgtac 159
|||||
Db 1 CAAGACACATTTATTACTGTAGAACACTGCTTATGTGTGTAC 51

RESULT 9

AX160577 51 bp DNA linear PAT 22-JUN-2001
LOCUS Sequence 3905 from Patent WO0140521.
AX160577
DEFINITION AX160577
VERSION AX160577.1 GI:14541908
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 51)
AUTHORS Shimkets,R.A. and Leach,M.
TITLE Nucleic acids containing single nucleotide polymorphisms and methods of use thereof
JOURNAL Patent: WO 0140521-A 3905 07-JUN-2001;
Curagen Corporation (US)

FEATURES
source Location/Qualifiers
1..51
misc_feature /organism="Homo sapiens"
26 /db_xref="taxon:9606"
Accession number cg43924289"

BASE COUNT 20 a 8 c 8 g 15 t
ORIGIN

Query Match 8.6%; Score 51; DB 6; Length 51;
Best Local Similarity 100.0%; Pred. No. 3.5e-16;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 223 taaagtgtcgcagttcttaaaacaccttcttcaatatgcatg 273
|||||
Db 1 TAAAGTGTGTCAGTAAAAAACACACCTTTCTTCATATATGCGCATG 51

RESULT 10

AX160578 50 bp DNA linear PAT 22-JUN-2001
LOCUS Sequence 3906 from Patent WO0140521.
AX160578
DEFINITION AX160578
VERSION AX160578.1 GI:14541909
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 50)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE Nucleic acids containing single nucleotide polymorphisms and methods of use thereof
JOURNAL Patent: WO 0140521-A 3906 07-JUN-2001;
Curagen Corporation (US)

FEATURES
source Location/Qualifiers
1..50
misc_feature /organism="Homo sapiens"
25..26 /db_xref="taxon:9606"
/note="Nucleotide deleted between bases 25 and 26
Accession number cg43924289"

BASE COUNT 19 a 8 c 8 g 15 t
ORIGIN

Query Match 5.7%; Score 34; DB 6; Length 50;
Best Local Similarity 100.0%; Pred. No. 5.2e-07;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 240 aaaaaaaaaaaccttcttcaatatgcatg 273
|||||
Db 17 AAAAAAAAAACACCTTTCTTCATATATGCGCATG 50

RESULT 11

AX160576 51 bp DNA linear PAT 22-JUN-2001
LOCUS Sequence 3904 from Patent WO0140521.
AX160576
DEFINITION AX160576
VERSION AX160576.1 GI:14541907
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 51)
AUTHORS Shimkets,R.A. and Leach,M.
TITLE Nucleic acids containing single nucleotide polymorphisms and methods of use thereof
JOURNAL Patent: WO 0140521-A 3904 07-JUN-2001;
Curagen Corporation (US)

FEATURES
source Location/Qualifiers
1..51
misc_feature /organism="Homo sapiens"
26 /db_xref="taxon:9606"
Accession number cg43924289"

BASE COUNT 14 a 8 c 9 g 20 t
ORIGIN

Query Match 4.2%; Score 25; DB 6; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.037;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 109 caagacacatttactttagaaccttcttattgtgtgtac 133
|||||
Db 1 CAAGACACATTTATTACTGTAG 25

RESULT 12

AX160574 50 bp DNA linear PAT 22-JUN-2001
LOCUS Sequence 3902 from Patent WO0140521.
AX160574
DEFINITION AX160574
VERSION AX160574.1 GI:14541905
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 50)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE Nucleic acids containing single nucleotide polymorphisms and methods of use thereof
JOURNAL Patent: WO 0140521-A 3902 07-JUN-2001;
Curagen Corporation (US)

FEATURES
source Location/Qualifiers
1..50
misc_feature /organism="Homo sapiens"
25..26 /db_xref="taxon:9606"
/note="Nucleotide deleted between bases 25 and 26
Accession number cg43924289"

BASE COUNT 12 a 8 c 3 g 27 t
ORIGIN

ORIGIN

Query Match 4.0%; Score 24; DB 6; Length 50;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 acatctatgttaccagcttct 24
|||||
Db 27 ACATCTATGTTTACAGGCTTCT 50

RESULT 13

AX160573 51 bp DNA linear PAT 22-JUN-2001
LOCUS Sequence 3901 from Patent W00140521.
DEFINITION AX160573
ACCESSION AX160573
VERSION AX160573.1 GI:14541904
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 51)
Nucleic acids containing single nucleotide polymorphisms and
methods of use thereof
JOURNAL Patent: WO 0140521-A 3901 07-JUN-2001;
Curagen Corporation (US)

FEATURES

source
1..51
Location/Qualifiers
misc_feature
26
/note="1 of 2 allelic variants (3902 is other entry)"
/db_xref="taxon:9606"
Accession number G943924289"
BASE COUNT 12 a 8 c 3 g 28 t
ORIGIN

Query Match 4.0%; Score 24; DB 6; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 acatctatgttaccagcttct 24
|||||
Db 28 ACATCTATGTTTACAGGCTTCT 51

RESULT 14

AL670276 149128 bp DNA linear HTG 30-JAN-2002
LOCUS Mus musculus chromosome 4 clone RP23-115D21, *** SEQUENCING IN
DEFINITION
PROGRESS *** in unordered pieces.

ACCESSION AL670276
VERSION AL670276.2 GI:18477095
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (sites)

REFERENCE
AUTHORS Burdon, J.
TITLE Direct Submission
JOURNAL Submitted (23-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Feb 1, 2002 this sequence version replaced gi:18307375.

COMMENT

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk

Project Information

Center project name: BM15D21
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 146474 bases at least Q40
Consensus quality: 146902 bases at least Q30
Consensus quality: 147368 bases at least Q20
Insert size: 148128; sum-of-contigs
Insert size: 154153; 3.3% error; agarose-fp
Quality coverage: 10.26x in Q20 bases; sum-of-contigs Quality
coverage: 9.86x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES

source
Location/Qualifiers
1..149128
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-115D21"
/clone_1tb="RPCT-23"
1..4701
/note="assembly-fragment:02491
fragment_chain:1
clone_end:SP6
vector_side:left"
4802..9220
/note="assembly-fragment:00966
fragment_chain:1"
9321..12084
/note="assembly-fragment:00594
fragment_chain:1"
12185..17776
/note="assembly-fragment:01529
fragment_chain:1"
17877..22373
/note="assembly-fragment:01073
fragment_chain:1"
22474..30439
/note="assembly-fragment:00975
fragment_chain:1"
30540..37808
/note="assembly-fragment:00276
fragment_chain:1"
37909..117226
/note="assembly-fragment:01606
fragment_chain:1"
117327..133040
/note="assembly-fragment:01524
fragment_chain:1"
133141..139559
/note="assembly-fragment:00876
fragment_chain:1"
139660..149128
/note="assembly-fragment:01484
clone_end:T7
vector_side:right"
BASE COUNT 39357 a 33291 c 33024 g 42447 t 1009 others
ORIGIN

Query Match 3.9%; Score 23; DB 2; Length 149128;
Best Local Similarity 100.0%; Pred. No. 0.42;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 239 aaaaaaaaaaacacttcttt 261
|||||
Db 147645 AAAAAAAAAACACCTTTCTTT 147667

```

RESULT 15
AC092096
LOCUS
DEFINITION Mus musculus chromosome 10 clone rp23-39k4, WORKING DRAFT SEQUENCE,
19 unordered pieces.
ACCESSION AC092096
VERSION AC092096.17 GI:18497142
KEYWORDS HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 336028)
AUTHORS Do,T. and Roe,B.A.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 336028)
AUTHORS Do,T. and Roe,B.A.
JOURNAL Direct Submission
Submitted (20-JUN-2001) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
On Feb 5, 2002 this sequence version replaced gi:18201826.
-----
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
3074: contig of 3074 bp in length
3075 3174: gap of unknown length
3175 6014: contig of 2840 bp in length
6015 6114: gap of unknown length
6115 9345: contig of 3231 bp in length
9346 9445: gap of unknown length
9446 12107: contig of 2662 bp in length
12108 12207: gap of unknown length
12208 18756: contig of 6549 bp in length
18757 18856: gap of unknown length
18857 23733: contig of 4877 bp in length
23734 23833: gap of unknown length
23833 30446: contig of 6613 bp in length
30447 30546: gap of unknown length
30547 37867: contig of 7321 bp in length
37868 37967: gap of unknown length
37968 45708: contig of 7741 bp in length
45709 45809: gap of unknown length
45809 53387: contig of 7579 bp in length
53388 53487: gap of unknown length
53488 63026: contig of 9539 bp in length
63027 63126: gap of unknown length
63127 74700: contig of 11574 bp in length
74701 74800: gap of unknown length
74801 85536: contig of 10736 bp in length
85537 85636: gap of unknown length
85637 108073: contig of 22437 bp in length
108074 108173: gap of unknown length
108174 137882: contig of 29709 bp in length
137883 137982: gap of unknown length
137983 183433: contig of 45451 bp in length
183434 183533: gap of unknown length
183534 222828: contig of 39295 bp in length
222829 262269: gap of unknown length
262270 336028: contig of 39341 bp in length
336028 336028: gap of unknown length
336028 336028: contig of 73659 bp in length.

```

```

FEATURES
    source
        location/Qualifiers
            1..336028
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /chromosome="10"
            /clone="rp23-39k4"
            /clone_1b="rp23"
BASE COUNT 101266 a 67078 c 66482 g 99351 t 1851 others
ORIGIN
Query Match 3.9% Score 23; DB 2; Length 336028;
Best Local Similarity 100.0%; Pred. No. 0.42;
Matches 23; Conservative 0; Mismatches 0; Gaps 0;
Qy 239 aaaaaaaaaaaccttcttctt 261
| | | | | | | | | | | | | | | |
Db 62506 AAAAAAAAAAACACCTTTCTTT 62528

```

Search completed: May 22, 2002, 08:50:10
 Job time: 11900 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 08:36:50 ; Search time 373 Seconds

(without alignments)
2729.572 Million cell updates/sec

Title: US-09-693-205-7_COPY_12200_12792

Perfect score: 593
Sequence: 1 acatctatgtttacaggtc.....acctccacatttatgtctt 593

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Genesed_032802.*

1: /SIDS1/gcgcdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS1/gcgcdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS1/gcgcdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS1/gcgcdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS1/gcgcdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS1/gcgcdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS1/gcgcdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS1/gcgcdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS1/gcgcdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS1/gcgcdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS1/gcgcdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS1/gcgcdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS1/gcgcdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS1/gcgcdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS1/gcgcdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS1/gcgcdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS1/gcgcdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS1/gcgcdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS1/gcgcdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS1/gcgcdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS1/gcgcdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS1/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS1/gcgcdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS1/gcgcdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	593	100.0	1317	22	AAAS29058
2	593	100.0	12792	22	AAH20176
3	593	100.0	12793	22	AAH20174
4	593	100.0	12793	22	AAH20178
5	593	100.0	12793	22	AAH20179
6	593	100.0	12793	22	AAH20182
7	430	72.5	1387	22	AAAS29132
8	51	8.6	51	22	AAI76962
9	51	8.6	51	22	AAI76964

10	34	5.7	50	22	AAI76965	Human silent SNP c
11	25	4.2	51	22	AAI76963	Human silent SNP c
12	24	4.0	50	22	AAI76961	Human silent SNP c
13	24	4.0	51	22	AAI76960	Human silent SNP c
14	21	3.5	4604	22	AAAS45491	Chemically pretrea
15	20	3.4	268	20	AAV90383	EST clone DM420.
16	20	3.4	1947	22	AAH16703	Human CDNA sequenc
17	19	3.2	384	21	AAH30804	Human colon cancer
18	19	3.2	504	23	ABL26223	Drosophila melanog
19	19	3.2	2504	23	ABL26222	Drosophila melanog
20	19	3.2	2617	23	ABL21778	Drosophila melanog
21	19	3.2	5238	24	ABL32758	Human immune syste
22	19	3.2	6052	24	ABL32417	Human immune syste
23	19	3.2	6081	24	ABL33428	Human immune syste
24	19	3.2	8770	22	AAAS46571	Tumour suppressor
25	19	3.2	8770	24	AAAS61353	Human gene regulat
26	18	3.0	462	21	AAAS9032	Human secreted pro
27	18	3.0	462	22	ABA42960	Human breast cell
28	18	3.0	462	22	ABA53380	Human foetal liver
29	18	3.0	462	22	ABA23158	Probe #1624 for ge
30	18	3.0	462	22	AAK01647	Human brain expres
31	18	3.0	462	22	AAK27093	Human bone marrow
32	18	3.0	462	22	AAI11687	Probe #1620 for ge
33	18	3.0	462	22	AAI32989	Probe #1675 used t
34	18	3.0	462	22	AAI01613	Human CDNA clone (
35	18	3.0	570	22	AAH10393	CDNA encoding a hu
36	18	3.0	784	21	AAAS8994	CDNA encoding a hu
37	18	3.0	1026	22	AAAS28472	Genomic sequence #
38	18	3.0	1026	22	AAAS28472	Genomic sequence #
39	18	3.0	1026	22	AAAS28472	Genomic sequence #
40	18	3.0	1037	21	AAAC59721	Human secreted pro
41	18	3.0	1075	22	ABAI7718	Human nervous syst
42	18	3.0	1711	21	AACT7140	Human ORFX ORF2655
43	18	3.0	1770	22	AAH17519	Human CDNA sequenc
44	18	3.0	1972	22	AAAD07834	Human secreted pro
45	18	3.0	2573	18	AAAT77804	CDNA encoding vari

ALIGNMENTS

RESULT 1
ID AAS29058 standard; CDNA: 1317 BP.
XX AAS29058;
AC
XX
XX 21-NOV-2001 (first entry)
XX
XX CDNA encoding for human DNA-binding protein #29.
XX
DE Human: DNA-binding protein; histone; chromo domain protein;
XX chromatin organisation modifier; y-box binding protein;
XX DNA organisation; gene transcription; malignant disease;
XX autoimmune disorder; rheumatic disease; genetic abnormality;
XX infectious disease; neurological disorder; gene therapy;
XX immunomodulatory; anti-HIV; anti rheumatic; anti microbial;
XX cytostatic; ss.
XX
OS Homo sapiens.
XX
XX WO200155162-A1.
XX
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001MO-US01305.
XX
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.

PR	18-APR-2000	2000US-0199122
PR	19-MAY-2000	2000US-0205467
PR	28-JUN-2000	2000US-0215135
PR	30-JUN-2000	2000US-0215186
PR	07-JUL-2000	2000US-0216647
PR	07-JUL-2000	2000US-0216880
PR	11-JUL-2000	2000US-0217487
PR	11-JUL-2000	2000US-0217496
PR	14-JUL-2000	2000US-0218290
PR	26-JUL-2000	2000US-0220964
PR	14-AUG-2000	2000US-0224518
PR	14-AUG-2000	2000US-0224519
PR	14-AUG-2000	2000US-0225213
PR	14-AUG-2000	2000US-0225214
PR	14-AUG-2000	2000US-0225266
PR	14-AUG-2000	2000US-0225267
PR	14-AUG-2000	2000US-0225268
PR	14-AUG-2000	2000US-0225270
PR	14-AUG-2000	2000US-0225477
PR	14-AUG-2000	2000US-0225757
PR	14-AUG-2000	2000US-0225758
PR	14-AUG-2000	2000US-0225759
PR	18-AUG-2000	2000US-0226279
PR	22-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0227182
PR	23-AUG-2000	2000US-0227109
PR	30-AUG-2000	2000US-0228924
PR	01-SEP-2000	2000US-0228287
PR	01-SEP-2000	2000US-0228343
PR	01-SEP-2000	2000US-0228344
PR	01-SEP-2000	2000US-0228345
PR	05-SEP-2000	2000US-0228519
PR	05-SEP-2000	2000US-0228533
PR	06-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231242
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0233080
PR	12-SEP-2000	2000US-0233081
PR	14-SEP-2000	2000US-0233397
PR	14-SEP-2000	2000US-0233398
PR	14-SEP-2000	2000US-0233399
PR	14-SEP-2000	2000US-0233401
PR	14-SEP-2000	2000US-0233401
PR	25-SEP-2000	2000US-0234597
PR	25-SEP-2000	2000US-0234598
PR	26-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235836
PR	29-SEP-2000	2000US-0236327
PR	29-SEP-2000	2000US-0236327
PR	29-SEP-2000	2000US-0236368
PR	29-SEP-2000	2000US-0236368
PR	29-SEP-2000	2000US-0236370
PR	29-SEP-2000	2000US-0236370
PR	02-OCT-2000	2000US-0236802
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239935
PR	13-OCT-2000	2000US-0239937

PR	20-OCT-2000;	2000US-0240960.
PR	20-OCT-2000;	2000US-0241221.
PR	20-OCT-2000;	2000US-0241785.
PR	20-OCT-2000;	2000US-0241786.
PR	20-OCT-2000;	2000US-0241787.
PR	20-OCT-2000;	2000US-0241808.
PR	20-OCT-2000;	2000US-0241809.
PR	20-OCT-2000;	2000US-0241826.
PR	01-NOV-2000;	2000US-0244617.
PR	01-NOV-2000;	2000US-0244617.
PR	08-NOV-2000;	2000US-0246474.
PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249219.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
XX	Rosen CA, Barash SC, Ruben SM;	
XX		
DR	WPI: 2001-465557/50.	
DR	P-PSDB: AAU18182.	
XX		
XX	Nucleic acid molecules encoding human secreted chromosomal binding	
PT	proteins, used in preventing, treating or ameliorating a disorder, e.g.	
PT	Alzheimer's and Parkinson's diseases and cancers -	
XX		
PS	Claim 4: SEQ ID No 39; 561pp; English.	
XX		
CC	The present invention relates to the isolation of novel DNA-binding	
CC	proteins (AAU18154-AAU18281), and cDNA and genomic sequences encoding	

Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acatctatgtttacagagcttctctgtttgtatgaagatagcaagaaactcaaatggt 60
 |||||
 Db 12200 acatctatgtttacagagcttctctgtttgtatgaagatagcaagaaactcaaatggt 12259

QY 61 ggcagttcttattacacaggtttagtatgtttcttgcgaacgtcttgcgaagaacatt 120
 |||||
 Db 12260 ggcagttcttattacacaggtttagtatgtttcttgcgaacgtcttgcgaagaacatt 12319

QY 121 tattaactgttgaagaaactgtcttattgttgtgtgtacattttcccaaatgttata 180
 |||||
 Db 12320 tattaactgttgaagaaactgtcttattgttgtgtgtacattttcccaaatgttata 12379

QY 181 attatatagttgtgttgaacagatgcaatcttctgttctcaaaagtgtcgaattaa 240
 |||||
 Db 12380 attatatagttgtgttgaacagatgcaatcttctgttctcaaaagtgtcgaattaa 12439

QY 241 aaaaaaacaacaccttcttctcaatagcgatgtaagtgaggtttttaaacttaaaac 300
 |||||
 Db 12440 aaaaaaacaacaccttcttctcaatagcgatgtaagtgaggtttttaaacttaaaac 12499

QY 301 atcaaaaatgttcaaaaatcattgttattatctagtagtttaattatcgcttatattc 360
 |||||
 Db 12500 atcaaaaatgttcaaaaatcattgttattatctagtagtttaattatcgcttatattc 12559

QY 361 cccatgaatgatacagactacattcaattcatgttgcgcacatgcttcttactt 420
 |||||
 Db 12560 cccatgaatgatacagactacattcaattcatgttgcgcacatgcttcttactt 12619

QY 421 aacatattcttctgcagaatgtaaaagtaataatagttatataaagtgtactcg 480
 |||||
 Db 12620 aacatattcttctgcagaatgtaaaagtaataatagttatataaagtgtactcg 12679

QY 481 ctgtaaatgtcttaataacttatacttataaggaaggttacagaacatggtgaacct 540
 |||||
 Db 12680 ctgtaaatgtcttaataacttatacttataaggaaggttacagaacatggtgaacct 12739

QY 541 ttttactttatttgggaataagaaatgttgcacactccacatttattgtct 593
 |||||
 Db 12740 ttttactttatttgggaataagaaatgttgcacactccacatttattgtct 12792

RESULT 3
 AAH20174
 ID AAH20174 standard; DNA: 12793 BP.
 XX
 AC AAH20174:
 XX
 DT 09-AUG-2001 (first entry)
 XX
 DE Human spastin nucleotide sequence SEQ ID NO:1.
 XX
 KW Human: mouse; spastin; ARSACS; chromosome 13q11; identification;
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibres;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 77..11566
 FT /*tag= a
 FT /product= "spastin"
 XX
 PN WO200129266-A2.
 XX
 PD 26-APR-2001.
 XX
 PF 20-OCT-2000; 2000MO-US29130.
 XX

PR 20-OCT-1999; 9905-0160588.
 XX
 PA (UYMC-) UNIV MCGILL.
 PA (HOP1-) HOPITAL SAINTE-JUSTINE.
 XX
 PI Hudson TJ, Engert J, Richter A;
 XX
 DR WPI: 2001-308494/32.
 XX P-PSDB; AAB97819.
 XX
 PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -
 XX
 PS Claim 1; Fig 9; 76pp; English.
 XX
 CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSACS)) gene sequences (1). The spastin gene has been mapped to
 CC chromosome 13q11. (1) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (1), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (1) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (1) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (1). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with spastin
 CC protein or spastin gene expression or activity. The present sequence
 CC encodes human spastin as given in the present invention.
 XX
 SQ Sequence 12793 BP; 4163 A; 2256 C; 2487 G; 3887 T; 0 other;

Query Match 100.0%; Score 593; DB 22; Length 12793;
 Best Local Similarity 100.0%; Pred. No. 1.7e-272;
 Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acatctatgtttacagagcttctctgtttgtatgaagatagcaagaaactcaaatggt 60
 |||||
 Db 12201 acatctatgtttacagagcttctctgtttgtatgaagatagcaagaaactcaaatggt 12260

QY 61 ggcagttcttattacacaggtttagtatgtttcttgcgaacgtcttgcgaagaacatt 120
 |||||
 Db 12261 ggcagttcttattacacaggtttagtatgtttcttgcgaacgtcttgcgaagaacatt 12320

QY 121 tattaactgttgaagaaactgtcttattgttgtgtgtacattttcccaaatgttata 180
 |||||
 Db 12321 tattaactgttgaagaaactgtcttattgttgtgtgtacattttcccaaatgttata 12380

QY 181 attatatagttgtgttgaacagatgcaatcttctgttctcaaaagtgtcgaattaa 240
 |||||
 Db 12381 attatatagttgtgttgaacagatgcaatcttctgttctcaaaagtgtcgaattaa 12440

QY 241 aaaaaaacaacaccttcttctcaatagcgatgtaagtgaggtttttaaacttaaaac 300
 |||||
 Db 12441 aaaaaaacaacaccttcttctcaatagcgatgtaagtgaggtttttaaacttaaaac 12500

QY 301 atcaaaaatgttcaaaaatcattgttattatctagtagtttaattatcgcttatattc 360
 |||||
 Db 12501 atcaaaaatgttcaaaaatcattgttattatctagtagtttaattatcgcttatattc 12560

QY 361 cccatgaatgatacagactacattcaattcatgttgcgcacatgcttcttactt 420

Db 12561 cccatgaatgacgacgacgacattatcatcattgtctgcgcacgctcttactt 12620
 QY 421 aacatattcttctgcagaatgtaaaaggaatgataattagttatataagttactg 480
 Db 12621 aacatattcttctgcagaatgtaaaaggaatgataattagttatataagttactg 12680
 QY 481 ctgtaaatgctgtaaatatacttattgcaattaggctctacgaacatgttgaact 540
 Db 12681 ctgtaaatgctgtaaatatacttattgcaattaggctctacgaacatgttgaact 12740
 QY 541 ttttactttattggaataaagaatgttgcacccacatttttgcct 593
 Db 12741 ttttactttattggaataaagaatgttgcacccacatttttgcct 12793

RESULT 4
 AAH20178
 ID AAH20178 standard; DNA; 12793 BP.
 XX AAH20178:
 DT 09-AUG-2001 (first entry)
 XX
 DE Human mutated spastin nucleotide sequence SEQ ID NO:11.
 XX
 KW Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibres;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN MO200129266-A2.
 XX
 PD 26-APR-2001.
 XX
 PF 20-OCT-2000; 2000MO-US29130.
 XX
 PR 20-OCT-1999; 99US-0160588.
 XX
 PA (UYMC-) UNIV MCGILL.
 PA (HOP1-) HOPITAL SAINTE-JUSTINE.
 XX
 PI Hudson TJ, Engert J, Richter A;
 XX
 DR MPI; 2001-308494/32.
 XX
 PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -
 XX
 PS Claim 1; Page -: 76pp; English.
 XX
 CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for

CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with spastin
 CC protein or spastin gene expression or activity. The present sequence
 CC represents a mutated human spastin gene from the present invention.
 CC N.B. The present sequence is not given in the present specification but
 CC is derived from the human spastin nucleotide sequence (AAH20174) as
 CC stated on page 14.
 XX
 SQ Sequence 12793 BP; 4163 A; 2255 C; 2487 G; 3888 T; 0 other:

Query Match 100.0%; Score 593; DB 22; Length 12793;
 Best Local Similarity 100.0%; Pred. No. 1.7e-272;
 Matches 593; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1 acatctatgtttacagcttcctggtttgataagatagcaaggaacccaatgtgt 60
 Db 12201 acatctatgtttacagcttcctggtttgataagatagcaaggaacccaatgtgt 12260
 QY 61 ggcagctcttattacagcttggttattgttctcggaactgcttgccaagaacatt 120
 Db 12261 ggcagctcttattacagcttggttattgttctcggaactgcttgccaagaacatt 12320
 QY 121 tattaactgttagaacacttgcttattgttgtgtgtacatatattccacaatgttata 180
 Db 12321 tattaactgttagaacacttgcttattgttgtgtgtgtacatatattccacaatgttata 12380
 QY 181 attatatagtgtgtgtgaacagatgcaactttgttcttaagggtgtcagttaa 240
 Db 12381 attatatagtgtgtgtgaacagatgcaactttgttcttaagggtgtcagttaa 12440
 QY 241 aaaaaaaacaccccttcttccaatagcgatgagtgaggttttttaacttaaaac 300
 Db 12441 aaaaaaaacaccccttcttccaatagcgatgagtgaggttttttaacttaaaac 12500
 QY 301 atcaaaaattgttaaaatcaltgtgtactagtagttataattatcogcttatattc 360
 Db 12501 atcaaaaattgttaaaatcaltgtgtactagtagttataattatcogcttatattc 12560
 QY 361 cccatgaatgacgacgacgacattatcatatgttgcctgcacgctcttactt 420
 Db 12561 cccatgaatgacgacgacgacattatcatatgttgcctgcacgctcttactt 12620
 QY 421 aacatattcttctgcagaatgtaaaaggaatgataattagttatataagttactg 480
 Db 12621 aacatattcttctgcagaatgtaaaaggaatgataattagttatataagttactg 12680
 QY 481 ctgtaaatgctgtaaatatacttattgcaattaggctctacgaacatgttgaact 540
 Db 12681 ctgtaaatgctgtaaatatacttattgcaattaggctctacgaacatgttgaact 12740
 QY 541 ttttactttattggaataaagaatgttgcacccacatttttgcct 593
 Db 12741 ttttactttattggaataaagaatgttgcacccacatttttgcct 12793

RESULT 5
 AAH20179
 ID AAH20179 standard; DNA; 12793 BP.
 XX AAH20179:
 DT 09-AUG-2001 (first entry)
 XX
 DE Human mutated spastin nucleotide sequence SEQ ID NO:12.
 XX
 KW Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;

PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -
PS
PS Claim 1: Page -: 76pp: English.
XX
XX The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSACS)) gene sequences (1). The spastin gene has been mapped to
CC chromosome 13q11. (1) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (1), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (1) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (1) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (1). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with Spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes a mutated human spastin.
CC N.B. The present sequence is not given in the present specification but
CC is derived from the human spastin nucleotide sequence (AAH20174) as
CC stated on page 14.
XX
XX Sequence 12793 BP; 4163 A; 2257 C; 2487 G; 3886 T; 0 other;

Query Match 100.0%; Score 593; DB 22; Length 12793;
Best Local Similarity 100.0%; Pred. No. 1.7e-272;
Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acatctatgtttagcaggctctctgttgaagaagatgacaagcaagaaactcaaatggt 60
DB 12201 acatctatgtttagcaggctctctgttgaagaagatgacaagcaagaaactcaaatggt 12260
QY 61 ggcagttcttattaccagatgttgaatgttcttctggaacgtcttgccaagacaatt 120
DB 12261 ggcagttcttattaccagatgttgaatgttcttctggaacgtcttgccaagacaatt 12320
QY 121 tattaactgttagaacaactgcttattgttgtgtgacattttccacaagtgtata 180
DB 12321 tattaactgttagaacaactgcttattgttgtgtgacattttccacaagtgtata 12380
QY 181 attatatagttgtgttgaacagatgacaacttctgttcttcaaaagtgctcagttaa 240
DB 12381 attatatagttgtgttgaacagatgacaacttctgttcttcaaaagtgctcagttaa 12440
QY 241 aaaaaaaacaaccttcttccaatatgacatgtagtgaagtttcttaacttaaaac 300
DB 12441 aaaaaaaacaaccttcttccaatatgacatgtagtgaagtttcttaacttaaaac 12500
QY 301 atcaaaaatgtttaaaatcatctgttatactagttatttaatttggcttatttc 360
DB 12501 atcaaaaatgtttaaaatcatctgttatactagttatttaatttggcttatttc 12560
QY 361 cccatgaatgatcagaactgacattcaatcagttgttcgcacatgctcttcaatt 420
DB 12561 cccatgaatgatcagaactgacattcaatcagttgttcgcacatgctcttcaatt 12620
QY 421 aacataatcttcttgcagaatgttaaaagtgaaatgaataatagttatataagtgctcgg 480
DB 12621 aacataatcttcttgcagaatgttaaaagtgaaatgaataatagttatataagtgctcgg 12680
QY 481 ctgtaaatgtatgctaataatacttatgcatlaagggtcttacagaacatgttgaactt 540

DB 12681 ctgtaaatgtatgctaataatacttatgcatlaagggtcttacagaacatgttgaactt 12740
QY 541 ttttactttatttgggaataaggaatgtttgcacccctccacatttatgctt 593
DB 12741 ttttactttatttgggaataaggaatgtttgcacccctccacatttatgctt 12793
RESULT 7
ID AAS29132 standard; CDNA; 1387 BP.
XX
XX AAS29132:
AC
AC 21-NOV-2001 (first entry)
DT
DT
XX
XX cDNA encoding for human DNA-binding protein #103.
DE
XX
XX Human; DNA-binding protein; histone; chromo domain protein;
KW chromatin organisation modifier; r-box binding protein;
KW DNA organisation; gene transcription; malignant disease;
KW autoimmune disorder; rheumatic disease; genetic abnormality;
KW infectious disease; neurological disorder; gene therapy;
KW immunomodulatory; anti-HIV; anti rheumatic; anti microbial;
KW cytostatic; ss.
XX
XX Homo sapiens.
OS
OS WO200155162-A1.
PN
PN 02-AUG-2001.
PD
PD
XX
XX 17-JAN-2001; 2001MO-US01305.
PE
PE
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225470.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.

```

PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0234999.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.

```

```

PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249224.
PR 17-NOV-2000; 2000US-0249225.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 06-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM.
XX WPI; 2001-465557/50.
XX P-PSDB; AAU18256.
DR
DR
XX
PT Nucleic acid molecules encoding human secreted chromosomal binding
PT proteins, used in preventing, treating or ameliorating a disorder, e.g.
PT Alzheimer's and Parkinson's diseases and cancers -
XX
XX
Claim 4; SEQ ID NO 113; 561pp; English.
XX
XX
The present invention relates to the isolation of novel DNA-binding
proteins (AAU18154-AAU18281), and cDNA and genomic sequences encoding
for these proteins. DNA-binding proteins such as histones, chromo
(CC (chromatin organisation modifier) domain proteins, and Y-box binding
CC proteins may contribute to diseases resulting from aberrant DNA
CC organisation and/or gene transcription. The sequences of the invention
are useful in screening assays to identify antagonists and/or agonists
CC that may enhance or block activities mediated by DNA-binding proteins.
CC Blockers of DNA-binding proteins may be useful in treating disorders
CC such as malignant diseases (e.g. cancer), autoimmune disorders
CC (e.g. diabetes mellitus), rheumatic diseases (e.g. rheumatoid
CC arthritis), genetic abnormalities (e.g. cystic fibrosis), infectious
CC diseases (e.g. HIV) and neurological disorders (e.g. Alzheimer's
CC disease). The polynucleotide sequences of the invention may also be
CC used in gene therapy. AAS29030-AAS29157 represent cDNA sequences
CC encoding for novel DNA-binding proteins.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1387 BP; 494 A; 165 C; 238 G; 482 T; 8 other;

```

```

Query Match 72.5%; Score 430; DB 22; Length 1387;
Best Local Similarity 100.0%; Pred. No. 7,1e-195;
Matches 430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 164 ttccacaatgtataattatagtggtgagcagatgacatcttggctc 223
DB 782 ttccacaatgtataattatagtggtgagcagatgacatcttggctc 841
QY 224 aaaggtgctgcagttataaaacacaccttcttcaatgtgcatgtagtgagtt 283
DB 842 aaaggtgctgcagttataaaacacaccttcttcaatgtgcatgtagtgagtt 901

```

CC assays to detect and quantitate the presence of similar nucleic acids
CC in samples, and therefore which patients may be in need of restorative
CC therapy. The polypeptides encoded by (1) may be used as antigens in the
CC production of antibodies specific for polymorphic polypeptides. The
CC antibodies may also be used to down regulate expression and activity.
CC The antibodies may also be used as diagnostic agents for detecting the
CC presence of polymorphic polypeptides in samples.
xx
SQ Sequence 51 BP; 15 A; 8 C; 8 G; 20 T; 0 other;

Query Match: 8.6%; Score 51; DB 22; Length 51;
Best Local Similarity 100.0%; Pred. No. 1.9e-10;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 caagacacattattactgtctagaacacctgcttactgtttggtgtac 159
|||||
Db 1 caagcacacattattactgtctagaacacctgcttactgtttggtgtac 51
|||||

RESUTⁿ 9
AA176964 ID AA176964 standard; DNA; 51 BP.
XX
AC AA176964;
XX
DT 09-NOV-2001 (first entry)
XX
DE Human silent SNP containing nucleic acid SEQ:3905.
XX
KM Human; single nucleotide polymorphism; SNP; genome; gene therapy;
KM protein therapy; vaccine; probe; diagnostic assay; detection;
KM quantitation; restorative therapy; polymorphic; ds.
XX
OS Homo sapiens.
XX
PN WO200140521-A2.
XX
PD 07-JUN-2001.
XX
PF 30-NOV-2000; 2000WO-US32758.
XX
PR 30-NOV-1999; 99US-0168138.
PR 29-NOV-2000; 2000US-0726173.
XX
PA (CUBA-) CUBAGEN CORP.
XX
PI Shinkets RA, Leach M;
XX
DR WPI: 2001-356160/37.
XX
PT Polymorphic nucleic acid sequences, useful in genetic testing and
XX therapy -
PS
XX
PS Claim 1; Page 1246; 2653bp; English.
XX
CC AA173060 to AA178667 represent isolated human polymorphic polynucleotide
CC sequences (I), which contain single nucleotide polymorphisms (SNPs).
CC AAM33114 to AAM33329 represent peptides related to human polymorphic
CC polynucleotide sequences. The sequences can be used in gene and protein
CC therapy, and in vaccine production. (I) and the polypeptides encoded by
CC them may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate expression of polymorphic polypeptides.
CC For example, (I) may be used to treat disorders by rectifying mutations
CC or deletions in a patient's genome that affect the activity of
CC polypeptides by expressing inactive proteins or to supplement the
CC patients own production of polypeptide. Additionally, (I) and its
CC complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar nucleic acids
CC in samples, and therefore which patients may be in need of restorative
CC therapy. The polypeptides encoded by (I) may be used as antigens in the
CC production of antibodies specific for polymorphic polypeptides. The
CC antibodies may also be used to down regulate expression and activity.

CC The antibodies may also be used as diagnostic agents for detecting the
CC presence of polymorphic polypeptides in samples.
XX
SQ Sequence 51 BP; 20 A; 8 C; 8 G; 15 T; 0 other;

Query Match 8.6%; Score 51; DB 22; Length 51;
Best Local Similarity 100.0%; Pred. No. 1.9e-14;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 223 taaggtctgcagtttaaaaaaacacaccttcttccaatgcatg 273
Db 1 taaggtctgcagtttaaaaaaacacaccttcttccaatgcatg 51

RESULT 10

AAI76965
ID AAI76965 standard; DNA; 50 BP.

XX
AC AAI76965;

XX
DT 09-NOV-2001 (first entry)

DE Human silent SNP containing nucleic acid SEQ:3906.

XX
KM Human; single nucleotide polymorphism; SNP; genome; gene therapy;

KW protein therapy; vaccine; probe; diagnostic assay; detection;
KM quantitation; restorative therapy; polymorphic; ds.

XX
OS Homo sapiens.

XX
PN WO200140521-A2.

XX
PD 07-JUN-2001.

XX
PF 30-NOV-2000; 2000WO-US32758.

XX
PR 30-NOV-1999; 99US-0168138.

XX
PR 29-NOV-2000; 2000US-0726173.

XX
PA (CURA-) CURAGEN CORP.

XX
PI Shinkets RA, Leach M;

XX
DR WPI; 2001-356160/37.

XX
PT Polymorphic nucleic acid sequences, useful in genetic testing and
PT therapy -

PS Claim 1; Page 1247; 2653pp; English.

XX
AAI73060 to AAI79867 represent isolated human polymorphic polynucleotide
sequences (I), which contain single nucleotide polymorphisms (SNPs).

CC AAM53114 to AAM53329 represent peptides related to human polymorphic
polynucleotide sequences. The sequences can be used in gene and protein
therapy, and in vaccine production. (I) and the polypeptides encoded by
CC them may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate expression of polymorphic polypeptides.

CC For example, (I) may be used to treat disorders by rectifying mutations
or deletions in a patient's genome that affect the activity of
CC polypeptides by expressing inactive proteins or to supplement the
CC patients own production of polypeptide. Additionally, (I) and its
CC complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar nucleic acids
CC in samples, and therefore which patients may be in need of restorative
CC therapy. The polypeptides encoded by (I) may be used as antigens in the
CC production of antibodies specific for polymorphic polypeptides. The
CC antibodies may also be used to down regulate expression and activity.

CC The antibodies may also be used as diagnostic agents for detecting the
CC presence of polymorphic polypeptides in samples.
XX
SQ Sequence 50 BP; 19 A; 8 C; 8 G; 15 T; 0 other;

Query Match 5.7%; Score 34; DB 22; Length 50;
Best Local Similarity 100.0%; Pred. No. 2.4e-06;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 240 aaaaaaacacaccttcttccaatgcatg 273
Db 17 aaaaaaacacaccttcttccaatgcatg 50

RESULT 11

AAI76963
ID AAI76963 standard; DNA; 51 BP;

XX
AC AAI76963;

XX
DT 09-NOV-2001 (first entry)

DE Human silent SNP containing nucleic acid SEQ:3904.

XX
KM Human; single nucleotide polymorphism; SNP; genome; gene therapy;

KW protein therapy; vaccine; probe; diagnostic assay; detection;
KM quantitation; restorative therapy; polymorphic; ds.

XX
OS Homo sapiens.

XX
PN WO200140521-A2.

XX
PD 07-JUN-2001.

XX
PF 30-NOV-2000; 2000WO-US32758.

XX
PR 30-NOV-1999; 99US-0168138.

XX
PR 29-NOV-2000; 2000US-0726173.

XX
PA (CURA-) CURAGEN CORP.

XX
PI Shinkets RA, Leach M;

XX
DR WPI; 2001-356160/37.

XX
PT Polymorphic nucleic acid sequences, useful in genetic testing and
PT therapy -

XX
PS Claim 1; Page 1246; 2653pp; English.

XX
AAI73060 to AAI79867 represent isolated human polymorphic polynucleotide
sequences (I), which contain single nucleotide polymorphisms (SNPs).

CC AAM53114 to AAM53329 represent peptides related to human polymorphic
polynucleotide sequences. The sequences can be used in gene and protein
CC therapy, and in vaccine production. (I) and the polypeptides encoded by
CC them may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate expression of polymorphic polypeptides.

CC For example, (I) may be used to treat disorders by rectifying mutations
or deletions in a patient's genome that affect the activity of
CC polypeptides by expressing inactive proteins or to supplement the
CC patients own production of polypeptide. Additionally, (I) and its
CC complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar nucleic acids
CC in samples, and therefore which patients may be in need of restorative
CC therapy. The polypeptides encoded by (I) may be used as antigens in the
CC production of antibodies specific for polymorphic polypeptides. The
CC antibodies may also be used to down regulate expression and activity.

CC The antibodies may also be used as diagnostic agents for detecting the
CC presence of polymorphic polypeptides in samples.
XX
SQ Sequence 51 BP; 14 A; 8 C; 9 G; 20 T; 0 other;

Query Match 4.2%; Score 25; DB 22; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.045;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 caagacacatttacttgtag 133
|||||
Db 1 caagacacatttacttgtag 25

RESULT 12

AA176961
ID AA176961 standard; DNA; 50 BP.

AC AA176961;

DT 09-NOV-2001 (first entry)

DE Human silent SNP containing nucleic acid SEQ:3902.

XX Human; single nucleotide polymorphism; SNP; genome; gene therapy;
KW protein therapy; vaccine; probe; diagnostic assay; detection;
KW quantitation; restorative therapy; polymorphic; ds.
XX

OS Homo sapiens.

XX WO200140521-A2.

XX 07-JUN-2001.

PF 30-NOV-2000; 2000WO-US32758.

PR 30-NOV-1999; 99US-0168138.

PR 29-NOV-2000; 2000US-0726173.

XX (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach M;

DR WPI; 2001-356160/37.

PT Polymorphic nucleic acid sequences, useful in genetic testing and
PT therapy -
XX
PS Claim 1; Page 1245; 2653pp; English.

XX AA173060 to AA179867 represent isolated human polymorphic polynucleotide
CC sequences (I), which contain single nucleotide polymorphisms (SNPs).
CC AA173060 to AA173329 represent peptides related to human polymorphic
CC polynucleotide sequences. The sequences can be used in gene and protein
CC therapy, and in vaccine production. (I) and the polypeptides encoded by
CC them may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate expression of polymorphic polypeptides.
CC For example, (I) may be used to treat disorders by rectifying mutations
CC or deletions in a patient's genome that affect the activity of
CC polypeptides by expressing inactive proteins or to supplement the
CC patient's own production of polypeptide. Additionally, (I) and its
CC complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar nucleic acids
CC in samples, and therefore which patients may be in need of restorative
CC therapy. The polypeptides encoded by (I) may be used as antigens in the
CC production of antibodies specific for polymorphic polypeptides. The
CC antibodies may also be used to down regulate expression and activity.
CC The antibodies may also be used as diagnostic agents for detecting the
CC presence of polymorphic polypeptides in samples.
XX

XX Sequence 50 BP; 12 A; 8 C; 3 G; 27 T; 0 other;

Query Match 4.0%; Score 24; DB 22; Length 50;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acatctatgtttacagcttct 24
|||||
Db 27 acatctatgtttacagcttct 50

RESULT 13

AA176960
ID AA176960 standard; DNA; 51 BP.

AC AA176960;

DT 09-NOV-2001 (first entry)

DE Human silent SNP containing nucleic acid SEQ:3901.

XX Human; single nucleotide polymorphism; SNP; genome; gene therapy;
KW protein therapy; vaccine; probe; diagnostic assay; detection;
KW quantitation; restorative therapy; polymorphic; ds.
XX

OS Homo sapiens.

XX WO200140521-A2.

XX 07-JUN-2001.

PF 30-NOV-2000; 2000WO-US32758.

PR 30-NOV-1999; 99US-0168138.

PR 29-NOV-2000; 2000US-0726173.

XX (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach M;

DR WPI; 2001-356160/37.

PT Polymorphic nucleic acid sequences, useful in genetic testing and
PT therapy -
XX
PS Claim 1; Page 1245; 2653pp; English.

XX AA173060 to AA179867 represent isolated human polymorphic polynucleotide
CC sequences (I), which contain single nucleotide polymorphisms (SNPs).
CC AA173060 to AA173329 represent peptides related to human polymorphic
CC polynucleotide sequences. The sequences can be used in gene and protein
CC therapy, and in vaccine production. (I) and the polypeptides encoded by
CC them may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate expression of polymorphic polypeptides.
CC For example, (I) may be used to treat disorders by rectifying mutations
CC or deletions in a patient's genome that affect the activity of
CC polypeptides by expressing inactive proteins or to supplement the
CC patient's own production of polypeptide. Additionally, (I) and its
CC complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar nucleic acids
CC in samples, and therefore which patients may be in need of restorative
CC therapy. The polypeptides encoded by (I) may be used as antigens in the
CC production of antibodies specific for polymorphic polypeptides. The
CC antibodies may also be used to down regulate expression and activity.
CC The antibodies may also be used as diagnostic agents for detecting the
CC presence of polymorphic polypeptides in samples.
XX

XX Sequence 51 BP; 12 A; 8 C; 3 G; 28 T; 0 other;

Query Match 4.0%; Score 24; DB 22; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acatctatgtttacagcttct 24
|||||
Db 28 acatctatgtttacagcttct 51

RESULT 14

AA545491/C
ID AA545491 standard; DNA; 4604 BP.

XX AA545491;

```

XX 18-DEC-2001 (First entry)
DE Chemically pretreated complementary DNA associated with cell cycle #98.
XX
XX
XX Cell cycle; human; Cpg dinucleotide; cytosine methylation; HIV; aging;
KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;
KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
KW PCR primer.
XX
XX Homo sapiens.
XX
XX WO200168911-A2.
XX
XX 20-SEP-2001.
XX
XX 15-MAR-2001; 2001WO-EP02945.
XX
XX 15-MAR-2000; 2000DE-1013847.
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIC-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
PI
XX WPI; 2001-602751/68.
DR
XX
XX Designing primers and probes for analysing diseases associated with
PT cytosine methylation state e.g. arthritis; cancer, aging
PT arteriosclerosis comprising fragments of chemically modified genes
PT associated with cell cycle -
XX
XX Claim 1; SEQ ID No 196; 28pp; English.
PS
XX
XX Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
CC molecules associated with the cell cycle and specific PCR primers of the
CC invention. The sequences are useful for detecting the methylation state
CC of all Cpg dinucleotides in a sequence and therefore for analysing
CC associated diseases. By analysing cytosine methylations in the pretreated
CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC of existing diseases or the predisposition to specific diseases can be
CC ascertained. The parameters may be compared to another set of genetic
CC and/or epigenetic parameters, the differences serving as basis for
CC diagnosis and/or prognosis events which are disadvantageous to patients.
CC The sequences of the invention are useful for the diagnosis and therapy
CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC aging, glomerular disease, Lewy body disease, arthritis,
CC arteriosclerosis, solid tumours and cancers.
XX
XX Sequence 4604 BP; 1054 A; 181 C; 1144 G; 2225 T; 0 other;
SQ

```

```

Query Match 3.5%; Score 21; DB 22; Length 4604;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 242 aaaaacacaccccttcttc 262
   ||||||||||||||||
Db 2569 AAAAAAACACCTTTCTTTC 2549

```

```

RESULT 15
AAV90383
ID AAV90383 standard; cDNA; 268 BP.
XX
AC AAV90383;
XX
XX 15-FEB-1999 (first entry)
DT

```

```

XX EST clone DM420.
DE
XX
XX Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KW gene therapy; ss.
XX
XX Homo sapiens.
OS
XX WO9845436-A2.
XX
XX 15-OCT-1998.
XX
XX 10-APR-1998; 98WO-US06955.
XX
XX 10-APR-1997; 97US-0838821.
XX
XX (GEMY) GENETICS INST INC.
XX
XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
XX
XX WPI; 1999-070077/06.
DR
XX
XX New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries.
XX
XX Claim 1; Page 523; 618pp; English.
PS
XX
XX The present sequence represents a human expressed sequence tag (EST).
CC The polynucleotide, which is a secreted EST, and the encoded protein
CC are predicted to have useful biological activities which would make
CC them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data is
CC given. Suggested activities include nutritional activity, immune
CC stimulating or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC activity, receptor/ligand activity, anti-inflammatory activity,
CC cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The polynucleotide may also be useful for gene therapy.
XX
XX Sequence 268 BP; 92 A; 47 C; 45 G; 84 T; 0 other;
SQ

```

```

Query Match 3.4%; Score 20; DB 20; Length 268;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 235 agttaaaaaaaaaaacact 254
   ||||||||||||||||
Db 176 agttaaaaaaaaaaacact 195

```

```

Search completed: May 22, 2002, 08:37:12
Job time: 6652 sec

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 08:31:08 ; Search time 91.58 Seconds

(without alignments)
1590.529 Million cell updates/sec

Title: US-09-693-205-7_COPY_12200_12792

Sequence: 1 acatcttctgtttacagcgtc.....acctccacattttgtctt 593

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters:

767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/ptodata1/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata1/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata1/1/ina/5A.COMB.seq: *
4: /cgn2_6/ptodata1/1/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata1/1/ina/6B.COMB.seq: *
6: /cgn2_6/ptodata1/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	18	3.0	6344	4 US-08-843-417-1	Sequence 1, Appl
C 2	17	2.9	164	4 US-09-172-711-56	Sequence 56, Appl
C 3	17	2.9	634	1 US-08-450-065-1	Sequence 1, Appl
C 4	17	2.9	634	1 US-08-450-595-1	Sequence 1, Appl
C 5	17	2.9	2659	1 US-08-007-775-3	Sequence 3, Appl
C 6	17	2.9	2848	2 US-08-805-918-1	Sequence 1, Appl
C 7	17	2.9	3417	2 US-08-464-402-1	Sequence 1, Appl
C 8	17	2.9	3417	4 US-09-054-775C-1	Sequence 1, Appl
C 9	17	2.9	4106	2 US-08-702-572-14	Sequence 14, Appl
C 10	17	2.9	4732	6 5521093-4	Patent No. 5521093
C 11	17	2.9	6038	4 US-09-305-639-4	Sequence 4, Appl
C 12	17	2.9	7622	4 US-09-305-639-1	Sequence 1, Appl
C 13	16	2.7	29	3 US-08-835-728D-96	Sequence 96, Appl
C 14	16	2.7	29	3 US-08-835-728D-200	Sequence 200, App
C 15	16	2.7	29	4 US-09-480-558-96	Sequence 96, App
C 16	16	2.7	29	4 US-09-480-558-200	Sequence 200, App
C 17	16	2.7	285	3 US-09-284-782-34	Sequence 34, Appl
C 18	16	2.7	285	3 US-09-284-782-35	Sequence 35, Appl
C 19	16	2.7	313	1 US-08-365-981-6	Sequence 6, Appl
C 20	16	2.7	633	1 US-08-234-783-1	Sequence 1, Appl
C 21	16	2.7	633	1 US-08-456-907-1	Sequence 1, Appl
C 22	16	2.7	633	5 PCT-US95-05523-1	Sequence 1, Appl
C 23	16	2.7	633	4 US-09-385-982-90	Sequence 90, Appl
C 24	16	2.7	675	1 US-08-307-499-55	Sequence 55, Appl
C 25	16	2.7	856	4 US-09-299-268-55	Sequence 55, Appl
C 26	16	2.7	856	1 US-08-117-373-10	Sequence 10, Appl
C 27	16	2.7	1065	3 US-08-591-685-6	Sequence 6, Appl

C 28	16	2.7	1288	1 US-08-047-041A-24	Sequence 24, Appl
C 29	16	2.7	1313	1 US-08-446-925-6	Sequence 6, Appl
C 30	16	2.7	1313	2 US-09-146-331-6	Sequence 6, Appl
C 31	16	2.7	1313	2 US-08-896-885-6	Sequence 6, Appl
C 32	16	2.7	1313	4 US-09-375-256-6	Sequence 6, Appl
C 33	16	2.7	1316	1 US-08-047-041A-11	Sequence 11, Appl
C 34	16	2.7	1316	2 US-08-795-006A-31	Sequence 31, Appl
C 35	16	2.7	1316	4 US-08-184-073-31	Sequence 31, Appl
C 36	16	2.7	1357	3 US-09-043-123-3	Sequence 3, Appl
C 37	16	2.7	1557	5 PCT-US96-05800-6	Sequence 6, Appl
C 38	16	2.7	1670	3 US-09-026-482B-1	Sequence 1, Appl
C 39	16	2.7	1684	2 US-08-899-811-21	Sequence 21, Appl
C 40	16	2.7	2504	1 US-08-484-105-15	Sequence 15, Appl
C 41	16	2.7	2504	1 US-08-484-106-15	Sequence 15, Appl
C 42	16	2.7	3480	1 US-07-657-769B-68	Sequence 68, Appl
C 43	16	2.7	3480	1 US-07-789-184-219	Sequence 219, App
C 44	16	2.7	3480	1 US-08-475-263-219	Sequence 219, App
C 45	16	2.7	3480	1 US-08-485-686-219	Sequence 219, App

ALIGNMENTS

RESULT 1
US-08-843-417-1/C
Sequence 1, Application US/08843417
Patent No. 6184349
GENERAL INFORMATION:
APPLICANT: Herman, Ronald C
APPLICANT: Delgado, Stephen G
APPLICANT: Fish, Linda M
APPLICANT: Sangameswaran, Lakshmi
APPLICANT: Rabert, Douglas K
TITLE OF INVENTION: CLONED PERIPHERAL NERVE
TITLE OF INVENTION: TETRODOXIN-RESISTANT SODIUM CHANNEL alpha-SUBUNIT
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Helier Ehirman White & Mcauliffe
STREET: 525 University Ave
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.A.
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/843,417
FILING DATE: April 15, 1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Schmoesees, William
REGISTRATION NUMBER: 31,796
TELEPHONE/DOCKET NUMBER: 28340-P1
TELEPHONE: (415)-324-7041
TELEFAX: (415)-324-0638
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6344 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: rat
TISSUE TYPE: Dorsal root ganglia
CELL TYPE: Peripheral nerve
US-08-843-417-1

Query Match 3.0%; Score 18; DB 4; Length 6344;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 282 ttttttaacttaaaaa 299
Db 6332 TTTTAACTTAAAAA 6315

RESULT 2
US-09-172-711-56
; Sequence 56, Application US/09172711
; Patent No. 6160105
; GENERAL INFORMATION:
; APPLICANT: Cunningham, Mary Jane
; APPLICANT: Zweiger, Gary B.
; APPLICANT: Panzer, Scott R.
; APPLICANT: Seilhamer, Jeffrey J.
; TITLE OF INVENTION: MONITORING TOXICOLOGICAL RESPONSES
; FILE REFERENCE: PA-0011 US
; CURRENT APPLICATION NUMBER: US/09/172,711
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PERL Program
; SEQ ID NO 56
; LENGTH: 164
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 700625315H1
US-09-172-711-56

Query Match 2.9%; Score 17; DB 4; Length 164;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 538 ctttttactttatt 554
Db 140 ctttttactttatt 156

RESULT 3
US-08-450-065-1/C
; Sequence 1, Application US/08450065
; Patent No. 5798105
; GENERAL INFORMATION:
; APPLICANT: Schoenmakers, Johannes G
; APPLICANT: Konings, Rudolph NH
; APPLICANT: Moelans, Inge IMD
; TITLE OF INVENTION: No. 5798105el protein
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Smithkline Beecham Corporate Patents -US
; STREET: UW2220, Po Box 1539
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,065
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/949645
; FILING DATE: 04-DEC-1992

ATTORNEY/AGENT INFORMATION:
NAME: Jervis, Herbert H.
REGISTRATION NUMBER: 31,171
REFERENCE/DOCKET NUMBER: B2992
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-270-5065
TELEFAX: 215-270-5090

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 634 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO

ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: PLASMODIUM

STRAIN: FALCIPARUM
IMMEDIATE SOURCE:
CLONE: 16K

US-08-450-065-1

Query Match 2.9%; Score 17; DB 1; Length 634;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 281 gtttttaactttaaa 297
Db 462 GTTTTAACTTAAA 446

RESULT 4
US-08-450-595-1/C
; Sequence 1, Application US/08450595
; Patent No. 5798106
; GENERAL INFORMATION:
; APPLICANT: Schoenmakers, Johannes G
; APPLICANT: Konings, Rudolph NH
; APPLICANT: Moelans, Inge IMD
; TITLE OF INVENTION: No. 5798106el protein
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Smithkline Beecham Corporate Patents -US
; STREET: UW2220, Po Box 1539
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,595
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/949645
; FILING DATE: 04-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Jervis, Herbert H.
; REGISTRATION NUMBER: 31,171
; REFERENCE/DOCKET NUMBER: B2992
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-270-5065
; TELEFAX: 215-270-5090
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 634 base pairs
; TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: PLASMODIUM
STRAIN: FALCIPARUM
IMMEDIATE SOURCE:
CLONE: 16K
US-08-450-595-1

Query Match 2.9%: Score 17; DB 1; Length 634;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 281 gtcttttaacttaaa 297
|||||
Db 462 GTTTTAACTTAA 446

RESULT 5
US-08-007-775-3
Sequence 3, Application US/08007775
Patent No. 5340733
GENERAL INFORMATION:
APPLICANT: Takashi UENO et al.
TITLE OF INVENTION: MBOI RESTRICTION-MODIFICATION GENES
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/007,775
FILING DATE: 19930122
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8650
TELEFAX:
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2659 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
HYPOTHETICAL:
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM: Moraxella bovis
STRAIN: 10900
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPOTYPE:

TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
LIBRARY:
CLONE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION: 12
IDENTIFICATION METHOD:
OTHER INFORMATION: /note= "Inosine"
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-08-007-775-3

Query Match 2.9%: Score 17; DB 1; Length 2659;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 294 taaacatcaaaaatt 310
|||||
Db 1208 TAAACATCAAAAATT 1224

RESULT 6
US-08-805-918-1/c
Sequence 1, Application US/08805918
Patent No. 5885821
GENERAL INFORMATION:
APPLICANT: MAGOTA, Koji
APPLICANT: MASUDA, Toyofumi
APPLICANT: SUZUKI, Yuji
APPLICANT: YABUTA, Masayuki
TITLE OF INVENTION: PROCESS FOR PRODUCTION OF SECRETORY KEY2
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/805,918
FILING DATE: 04-MAR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-073217
FILING DATE: 04-MAR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-352580

FILING DATE: 16-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meuth, Donna M.
REGISTRATION NUMBER: 36,607
REFERENCE/DOCKET NUMBER: 001560-295
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-6620
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2848 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Saccharomyces cerevisiae
STRAIN: X2180-1B
FEATURE:
NAME/KEY: CDS
LOCATION: 170..2611
US-08-805-918-1

Query Match 2.9%; Score 17; DB 2; Length 2848;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 331 tagtagttataattat 347
Db 157 TAGTAGTTATAATTAT 141

RESULT 7
US-08-464-402-1
Sequence 1, Application US/08464402
Patent No. 5858705
GENERAL INFORMATION:
APPLICANT: WEI, ET AL.
TITLE OF INVENTION: Human DNA Ligase III
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,402
FILING DATE: June 5, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/03939
FILING DATE: 31 MAR 95
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-388
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3417 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE

TOPOLOGY: LINEAR
MOLECULE TYPE: cDNA
US-08-464-402-1

Query Match 2.9%; Score 17; DB 2; Length 3417;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 232 tgcagttaaaaaaaaa 248
Db 3394 TGCAGTTAAAAAAA 3410

RESULT 8
US-09-054-775C-1
Sequence 1, Application US/09054775C
Patent No. 6284504
GENERAL INFORMATION:
APPLICANT: Wei, Yang-Fei
Yu, Guo-Liang
Haseltine, William
TITLE OF INVENTION: Human DNA Ligase III
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/054,775C
FILING DATE: 03-Apr-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/464,402
FILING DATE: 05-JUN-1995
APPLICATION NUMBER: PCT/US95/03939
FILING DATE: 31-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Hoover, Kenley K.
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PFI61D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3417 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-054-775C-1

Query Match 2.9%; Score 17; DB 4; Length 3417;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 232 tgcagttaaaaaaaaa 248
Db 3394 TGCAGTTAAAAAAA 3410

RESULT 9

US-08-702-572-14/C
; Sequence 14, Application US/08702572
; Patent No. 5965386
; GENERAL INFORMATION:
; APPLICANT: Kerry-Williams, Sean M
; APPLICANT: Gilbert, Sarah C
; TITLE OF INVENTION: Yeast Strains and Modified Albumins
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Centeon L.L.C.
; STREET: 1020 First Avenue
; City: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406-1310
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Microsoft Word 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/702,572
; FILING DATE: 11-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: NO 95/23857
; FILING DATE: 1-MAR-1995
; APPLICATION NUMBER: GB 9404270.2
; FILING DATE: 5-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Naomi Biswas
; REGISTRATION NUMBER: 38,384
; REFERENCE/DOCKET NUMBER: CE0114 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610/878/4294
; TELEFAX: 610/878/4221
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4106 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; US-08-702-572-14

Query Match 2.9%; Score 17; DB 2; Length 4106;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 331 tagtagttataattat 347
Db 1316 TAGTAGTTATAATTAT 1300

RESULT 10
5521093-4/C
; Patent No. 5521093
; APPLICANT: LEMKINE, YVES;NGUYEN, MARTINE;ACHSTETTER, TILMAN
; TITLE OF INVENTION: YEAST VECTOR CODING FOR HETEROLOGOUS
; GENE FUSIONS LINKED VIA KEX2 CLEAVAGE SITE AND CODING FOR
; TRUNCATED KEX2 GENES
; NUMBER OF SEQUENCES: 8
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/393,025
; FILING DATE: 23-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 191,354
; FILING DATE: 07-FEB-1994

APPLICATION NUMBER: 26,121
; FILING DATE: 04-MAR-1993
; APPLICATION NUMBER: 500,885
; FILING DATE: 29-MAR-1990
; SEQ ID NO:4:
; LENGTH: 4732
5521093-4

Query Match 2.9%; Score 17; DB 6; Length 4732;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 331 tagtagttataattat 347
Db 1316 TAGTAGTTATAATTAT 1320

RESULT 11
US-09-305-639-4
; Sequence 4, Application US/09305639
; Patent No. 6200778
; GENERAL INFORMATION:
; APPLICANT: Treco, Douglas A.
; APPLICANT: Seiden, Richard F.
; TITLE OF INVENTION: GENOMIC SEQUENCES FOR PROTEIN PRODUCTION AND DELIVERY
; FILE REFERENCE: 07236/016001
; CURRENT APPLICATION NUMBER: US/09/305,639
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: 60/084,663
; EARLIER FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 6038
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-305-639-4

Query Match 2.9%; Score 17; DB 4; Length 6038;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 435 gcagaatgtaaaagta 451
Db 1446 gcagaatgtaaaagta 1462

RESULT 12
US-09-305-639-1
; Sequence 1, Application US/09305639
; Patent No. 6200778
; GENERAL INFORMATION:
; APPLICANT: Treco, Douglas A.
; APPLICANT: Heartlein, Michael W.
; APPLICANT: Seiden, Richard F.
; TITLE OF INVENTION: GENOMIC SEQUENCES FOR PROTEIN PRODUCTION AND DELIVERY
; FILE REFERENCE: 07236/016001
; CURRENT APPLICATION NUMBER: US/09/305,639
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: 60/084,663
; EARLIER FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 7622
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-305-639-1

Query Match 2.9%; Score 17; DB 4; Length 7622;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 435 gcgaatgtataagcta 451
DB 1446 gcgaatgtataagcta 1462

RESULT 13

US-08-835-728D-96/C
Sequence 96, Application US/08835728D
Patent No. 6017704
GENERAL INFORMATION:
APPLICANT: Herman, James G.
APPLICANT: Baylin, Stephen B.
TITLE OF INVENTION: Methylation Specific Detection
NUMBER OF SEQUENCES: 216
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/835,728D
FILING DATE: April 11, 1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/656,716
FILING DATE: June 03, 1996,
ATTORNEY/AGENT INFORMATION:
NAME: Halle, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07265/125001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
FAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 29 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-835-728D-96

Query Match 2.7%; Score 16; DB 3; Length 29;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 239 aaaaaaaaaaacact 254
DB 28 AAAAAAAAAAACACT 13

RESULT 14

US-08-835-728D-200
Sequence 200, Application US/08835728D
Patent No. 6017704
GENERAL INFORMATION:
APPLICANT: Herman, James G.
APPLICANT: Baylin, Stephen B.
TITLE OF INVENTION: Methylation Specific Detection
NUMBER OF SEQUENCES: 216
CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/835,728D
FILING DATE: April 11, 1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/656,716
FILING DATE: June 03, 1996,
ATTORNEY/AGENT INFORMATION:
NAME: Halle, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07265/125001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
FAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 200:
SEQUENCE CHARACTERISTICS:
LENGTH: 29 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-835-728D-200

Query Match 2.7%; Score 16; DB 3; Length 29;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 239 aaaaaaaaaaacact 254
DB 2 AAAAAAAAAAACACT 17

RESULT 15

US-09-490-558-96/C
Sequence 96, Application US/09490558
Patent No. 6265171
GENERAL INFORMATION:
APPLICANT: Herman, James G.
APPLICANT: Baylin, Stephen B.
TITLE OF INVENTION: Methylation Specific Detection
NUMBER OF SEQUENCES: 216
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/490,558
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/835,728
FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Haile, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07265/125001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ. ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 29 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-09-490-558-96

Query Match 2.7%; Score 16; DB 4; Length 29;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 239 aaaaaaaaaaacct 254
|||||
Db 28 AAAAAAAAAACAACT 13

Search completed: May 22, 2002, 08:31:22
Job time: 7007 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 07:30:35 ; Search time 2968.03 Seconds
(without alignments)
2696.635 Million cell updates/sec

Title: US-09-693-205-7_COPY_12200_12792

Sequence: 1 aacatcttcttaccagctc.....acctccacatttattgctt 593

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 674847542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estda:*
2: em_esthum:*
3: em_estlin:*
4: em_estlinu:*
5: em_estlov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inu:*
15: em_gss_pln:*
16: em_gss_vtl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	423	71.3	500	9	AI803488 tcl7902.x
C 2	419	70.7	543	9	AI932370
C 3	413	69.6	828	10	BM470780 AGENCOURT
C 4	382	64.4	660	10	BE890125 601513104
C 5	373	62.9	497	9	AM087745 x68f08.x
C 6	368	62.1	368	9	AA683013 x68f08.x
C 7	355	59.9	356	10	BE856736 a681b08.s
C 8	355	59.9	356	10	BF438152 7667f12.x
C 9	345	58.2	469	9	AI217518 qh20908.x
C 10	341	57.5	422	9	AI076834 o246d05.x
C 11	330	55.6	410	9	AA954825 o039d10.s
C 12	310	52.3	528	9	AA417817 zv04h08.t
C 13	290	48.9	416	9	AI49896 tn97d12.x
C 14	273	46.0	404	9	AA809783 nm66c04.s
C 15	240	40.5	456	9	AA417676 zv04d08.r
C 16	233	39.3	632	9	AM968633 ESR380709
C 17	218	36.8	407	9	AA481507 aa34c04.s

C 18	208	35.1	739	9	AL567149
C 19	206	34.7	351	9	AM262498
C 20	192.	32.4	441	10	N59442
C 21	172	29.0	349	9	AI561086
C 22	165	27.8	483	10	N48291
C 23	164	27.7	287	10	T17045
C 24	155	26.1	418	10	N46342
C 25	130	21.9	1016	10	BE886315
C 26	106	17.9	789	10	BE889418
C 27	79	13.3	840	10	BF693898
C 28	74	12.5	203	9	AA228047
C 29	74	12.5	1079	10	BM476997
C 30	70	11.8	727	10	BC619034
C 31	46	7.8	236	9	AA228008
C 32	41	6.9	712	9	AV716805
C 33	41	6.9	818	10	BI599024
C 34	32	5.4	162	10	BE894617
C 35	29	4.9	478	10	BI871763
C 36	28	4.7	677	10	BF576832
C 37	26	4.4	201	9	AV167991
C 38	26	4.4	230	9	BB297230
C 39	26	4.4	292	9	AI447763
C 40	26	4.4	302	9	AI865982
C 41	26	4.4	428	10	BE862199
C 42	26	4.4	460	9	AI506773
C 43	26	4.4	530	10	BM118416
C 44	26	4.4	543	9	AM535384
C 45	26	4.4	566	10	BM119091

ALIGNMENTS

RESULT 1	AI803488/c	500 bp	mrna	linear	EST 13-DEC-1999
LOCUS	tcl7902.x1	Soares_NHMPu_S1	Homo sapiens	CDNA clone	IMAGE:2064146
DEFINITION	3', mRNA sequence.				
ACCESSION	AI803488				
VERSION	AI803488.1	GI:5368882			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 500)				
TITLE	NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap .				
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index				
COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgaps-remail.nih.gov This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Insert Length: 1089 Std Error: 0.00 Seq primer: -400P from Gldco High quality sequence stop: 447.				

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2064146"
/tissue_type="Soares_NHMPu_S1"
/tissue="Pooled human melanocyte, fetal heart, and pregnant uterus"
/lab_host="DH10b"
/note="Organ: mixed (see below); Vector: pRT730-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NbHm, pregnant uterus NbHpu, and fetal heart NBH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization

reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

BASE COUNT

196 a 80 c 62 g 162 t

Query Match 71.3%; Score 423; DB 9; Length 500;
Best Local Similarity 100.0%; Pred. No. 2.6e-153;
Matches 423; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 aaatgataatataatagatgctgagcagatgcaatcttctgctgtaaaagt 230
|||||
DB 500 AATGTAATATTAATTAATGAGTGAACAGAGATGCAATCTTTGTCCTAAAGT 441
|||||
QY 231 ctgcagtaataaaaaaacacaccttcttcaatagcattgagtggtttttaa 290
|||||
DB 440 CTCGAGTTAAAAAACACCTTTCTTCATATGCGATGAGTGAGTTTAA 381
|||||
QY 291 ctttaaaacatcaaaaatgtaaaatcattgctatctagttatattatcgg 350
|||||
DB 380 CTTTAAAAACATCAAAATGTTAAATCATGTGTTATCTAGATTAAATATCG 321
|||||
QY 351 ctataatcccatgaatgacagacatgaatcaatcattgctgctgcagatg 410
|||||
DB 320 CTTATATATTCCTCATGATGATGCAACATTAATTAATGTTGTCGCGATGCT 261
|||||
QY 411 tcttacttaacatattcttctgacgaatgtaaaagttaataatgattata 470
|||||
DB 260 TCTTACTTTAAACATATTTCTTTGCGAATGTAAAGTAAATAGTTTATATA 201
|||||
QY 471 agctgactgctgtaaatgcttaaatattatgcaatgaagggtctacagaa 530
|||||
DB 200 AGGTACTGCTGTAATATGCTTAATTAATTAATTAATGAAGGCTTACGAA 141
|||||
QY 531 gtgaacatttcttacttattggaataaagaatgcttgcaccccatcttatt 590
|||||
DB 140 GTTGAACCTTTTACTTTATTTGGGAATTAAGAAATGTTGACCTCCACATTTT 81
|||||
QY 591 ctt 593
|||||
DB 80 CTT 78

RESULT 2
AI932370 543 bp mRNA linear EST 17-DEC-1999
LOCUS

DEFINITION wd27e11.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2329388 3', mRNA sequence.
AI932370

ACCESSION AI932370.1 GI:5671107

VERSION

KEYWORDS

SOURCE

ORGANISM

human.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

This clone is available royalty-free through LNL; contact the
IMAG Consortium (info@image.llnl.gov) for further information.
Insert Length: 982 Std Error: 0.00
Seq primer: -40UP from Gdbco
High quality sequence stop: 455.
Location/Qualifiers
1.543
/organism="Homo sapiens"
/db_xref="taxon:9606"

/clone="IMAGE:2329388"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pRT3D-Pac (Pharmacia) with
a modified polylinker. Site1: Not I; Site2: Eco RI.
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI-GAP-GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."

Query Match 70.7%; Score 419; DB 9; Length 543;
Best Local Similarity 100.0%; Pred. No. 8.7e-152;
Matches 419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 actgttagaacactgtcttattgctgctgacataattccacaagtataatta 185
|||||
DB 543 ACTGTAGAACACTGCTTATGTTGTGTGATCATATTTCCACAAATGTTAATTTA 484
|||||
QY 186 tatagtgtgttgaaagagtgcaatcttctgtctaaagtgctgcagtaaaaaa 245
|||||
DB 483 TATAGTGTGTGTGAACGAGATGCAATCTTTGTGCTTAAAGGCTGCGAGTTAAAAA 424
|||||
QY 246 aacaacaccttcttcaatatgcatgtagtgagttcttcaacttaaaacatcaa 305
|||||
DB 423 AAACACCTTTCTTAAATATGCAATGATGAGTGTGATTTTAACTTAAAAACATCA 364
|||||
QY 306 aaattgttaaatcatgtgtatcctagtagttataatcatoggttataattcc 365
|||||
DB 363 AAATGTTAAATCATGTGTATCTAGTATTAATTAATGCGCTTATTTCCCAT 304
|||||
QY 366 gaatgacagaactgcaatttaattcattgtgtctgcagtgcttcttaacttaac 425
|||||
DB 303 GAATGATCAAGACATGCAATTAATTCATGTTGTCTGCGCATGCTTCTTAACAT 244
|||||
QY 426 attctcttgcaaatgtaaaagtaagtaattatgattataaagtgcagtcgtgta 485
|||||
DB 243 ATTTCCTTTCAGAGATTAAGATATGATTAATTAATTAATTAATTAATTAATTA 184
|||||
QY 486 aatgataatataatattatgcaattaaagggttaacagacatgttgaactttt 544
|||||
DB 183 AATGATGCTAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 125
|||||

RESULT 3

BM470780

LOCUS

DEFINITION

AGENCOURT_6463084 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5533575

5', mRNA sequence.
BM470780

ACCESSION BM470780.1 GI:18519822

VERSION

KEYWORDS

SOURCE

ORGANISM

human.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

This clone is available royalty-free through LNL; contact the
NIH-MGC http://mgc.ncl.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov
 Plate: LLAM1218 row: k column: 16
 High quality sequence stop: 680.

FEATURES

source

Location/Qualifiers
 1..828
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:553575"
 /clone_1lb="NIH_MGC_71"
 /tissue_type="leiomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: uterus; Vector: pCMV-Sport6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 2.1 kb."
 BASE COUNT 262 a 107 c 138 g 321 t
 ORIGIN

Query Match

Best Local Similarity 69.6%; Score 413; DB 10; Length 828;
 Best Local Similarity 99.7%; Pred. No. 1.4e-149;

Matches 593; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 acatctatggttacaaggtcctctgtttgatgaagatagcaaggaactcaaatggt 60
 |||
 DB 143 ACATCTTATGTTTACAGGCTTCTGTTGATGAGATGACAGGAAACCTCAAAATGCT 202
 QY 61 ggcagttctattacacagttgttagtattgtcttgcgaactgcttccagaacaatt 120
 |||
 DB 203 GGCAGTCTTATACCAAGTTGATGATGTTCTTGGAACCTGCTGCCAGCAACATT 262
 QY 121 tattaactgttagaacaactgtcttattgttgggtacataatttccacaatggtata 180
 |||
 DB 263 TATTACTGTAGAACACTGCTTATGTTGATGATGATGATGATGATGATGATGATGAT 322
 QY 181 attatagtggtgttggaacaggaatgcaacttcttggcttcaaggtgctgcaagt-- 238
 |||
 DB 323 ATTTATATAGTGTGTTGAGACAGATGCAATCTTTGTTCTTAAAGTCTGCGAGTTAA 382
 QY 239 aaaaaaaacaacacttcttccaatatgcatgtatgagatttttcaacttataa 298
 |||
 DB 383 AAAAAAACAACCTTTCTTCAATATGCGATGATGATGATGATGATGATGATGATGATGAT 442
 QY 299 acatcaaaaatgtttaaataatcattgttctatctagtagttataatcgcgcttata 358
 |||
 DB 443 ACATCAAAAAATGTTAAATCATGTTGATGATGATGATGATGATGATGATGATGATGAT 502
 QY 359 tccccatgatatgcatgaactgacattatcattatgttctcgcgcacatgcttact 418
 |||
 DB 503 TCCCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 562
 QY 419 ttaacatattctcttgcagaatgttaaaggtatgataatgattatataagttact 478
 |||
 DB 563 TTAACTATTTCTTTTGCAGAAATGTAAGGTAATGTAATGTAATGTAATGTAATGTAATG 622
 QY 479 ggcgtgaatgcatgaataatacttaataatgcaatgaaggtctacagaacatgttgaac 538
 |||
 DB 623 GCGTGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 682
 QY 539 ttttttacttattatgggaataaggaatgttgcacatccacatttattgctt 593
 |||
 DB 683 TTTTCTTACTTTTATTTGGAATGATGATGATGATGATGATGATGATGATGATGATGAT 737

RESULT 4

BE890125 BE890125 660 bp mRNA linear EST 20-OCT-2000
 LOCUS 601513104F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3914521 5',
 DEFINITION MRA sequence.
 ACCESSION BE890125
 VERSION BE890125.1 GI:10348134

KEYWORDS

EST.
 human.

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 660)

AUTHORS

NIH-MGC http://imgc.ncl.nih.gov/.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: c9apbs-remail.nih.gov

FEATURES

Location/Qualifiers
 1..660
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3914521"
 /clone_1lb="NIH_MGC_71"
 /tissue_type="leiomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: uterus; Vector: pCMV-Sport6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 2.1 kb."
 BASE COUNT 196 a 95 c 107 g 262 t
 ORIGIN

Query Match

Best Local Similarity 64.4%; Score 382; DB 10; Length 660;
 Best Local Similarity 99.6%; Pred. No. 1.5e-137;

Matches 552; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 acatctatggttacaaggtcctctgtttgatgaagatagcaaggaactcaaatggt 60
 |||
 DB 41 ACATCTTATGTTTACAGGCTTCTGTTGATGAGATGACAGGAAACCTCAAAATGCT 100
 QY 61 ggcagttctattacacagttgttagtattgtcttgcgaactgcttccagaacaatt 120
 |||
 DB 101 GGCAGTCTTATACCAAGTTGATGATGATGATGATGATGATGATGATGATGATGAT 160
 QY 121 tattaactgttagaacaactgtcttattgttgggtacataatttccacaatggtata 180
 |||
 DB 161 TATTACTGTAGAACACTGCTTATGTTGATGATGATGATGATGATGATGATGATGATGAT 220
 QY 181 attatagtggtgttggaacaggaatgcaacttcttggcttcaaggtgctgcaagt-a 239
 |||
 DB 221 ATTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 280
 QY 240 aaaaaaaacaacacttcttccaatatgcatgtatgagatttttcaacttataa 299
 |||
 DB 281 AAAAAAACAACCTTTCTTCAATATGCGATGATGATGATGATGATGATGATGATGATGAT 340
 QY 300 catcaaaaatgtttaaataatcattgttattcattagttatataatcgcgcttata 359
 |||
 DB 341 CATCAAAAATGTTAAATCATGTTGATGATGATGATGATGATGATGATGATGATGATGAT 400
 QY 360 ccccatgaatgcatgaactgacattcaatcattgttctcgcgcacatgcttactt 419
 |||
 DB 401 CCCCATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 460
 QY 420 taacatattcttctgcagaatgttaaaggtatgataatgattatataagttact 479
 |||
 DB 461 TAACTATTTCTTTTGCAGAAATGTAAGGTAATGTAATGTAATGTAATGTAATGTAATG 520
 QY 480 gctgtaaatgcatgaataatactttagcaatgaaggtctacagaacatgttgaact 539
 |||

```

Db 521 GGTGTAATGATGCTAAATTACTTTATGCAATTAAAGGGCTTACAGACATGTTGAAC 580
QY 540 tttttactttat 553
Db 581 TTTTTCATTCTTTAT 594

RESULT 5
AM087745/c 497 bp mRNA linear EST 15-OCT-1999
LOCUS xB68f08.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:2581479.3', mRNA sequence.
ACCESSION AM087745
VERSION AM087745
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
REFERENCE 1 (bases 1 to 497)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 455.
Location/Qualifiers
FEATURES
source
1..497
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2581479"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site: 1: Not I; Site 2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI-CGAP-GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 195 a 80 c 63 g 159 t
ORIGIN
Query Match 62.1%; Score 373; DB 9; Length 497;
Best Local Similarity 100.0%; Pred. No. 5.2e-134;
Matches 373; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 aatgtatataattatataatgtgtgaacagatcaatctttgtgtctaaagtgc 231
Db 497 AAGTTATATATTTATATATGAGTGGTTGAACAGATCAATCTTTGTGTAAAGGTGC 438
QY 232 tgcagtaaaaaaaacaaacaccttctctcaatgacatgagtgagttctttaa 291
Db 437 TGCAGTTAAAAAAAACAAACACCTTTCTTTCATATGCGATGAGTGAAGTCTTTTAAAC 378
QY 292 tttaaaacacaaataatgttaaatcatgtgtatcctagtagttataatcgcgc 351
Db 377 TTTAAAAACATCAAAATTTGTTAAATCATTTGTTATCTAGTAAATTATATCGGC 318
QY 352 ttatattcccatgatgatcagaactgacattatcatgttgcgcgcagctt 411
Db 317 TTTATATTTCCCATGATGATGCAAGCTGACATTTAATTCATGTTGTCTCGCCATGCTT 258

```

```

QY 412 cttacttiacaatattcttctgcagaatgtaaaagtaagataatagttatataa 471
Db 257 CTTTACTTTACADATTTCTTTTGCAGATGTAAGGCTAATGATATGTTATATTA 198
QY 472 gtgtactgctgtaaatgatgatataatcattatgcaatgaaggcttaagaatc 531
Db 197 GTGTACTGCTGCTTAATGATGATGATTAATGCAATTAAAGGGCTTACAGACATG 138
QY 532 ttgaacttttt 544
Db 137 TTGAACCTTTT 125

RESULT 6
AA683013/c 368 bp mRNA linear EST 15-DEC-1997
LOCUS ae81b08.s1 Stratagene schizo brain S11 Homo sapiens cDNA clone
DEFINITION IMAGE:970551.3', mRNA sequence.
ACCESSION AA683013
VERSION AA683013
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
REFERENCE 1 (bases 1 to 368)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Krizman, D., Kucab, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin,
J., Moore, B., Schellenger, K., Steptoe, M., Tan, F., Theising, B.,
White, Y., Wylie, T., Waterston, R., and Wilson, R.
WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson R.
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: polyT not found
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 358.
Location/Qualifiers
FEATURES
source
1..368
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:970551"
/clone_lib="Stratagene schizo brain S11"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: Bluescript SK-; Site: 1: EcoRI; Library
constructed from S-11 frontal lobe, male, 34 years old,
50% caucasian, 50% Aleutian. Schizophrenic suicide.
Random primed into EcoRI site of ZAP II Vector. Mass
excised. Custom library. Avg insert length 1.4kb.
Material obtained by Johnston N., Torrey, E.F., Yolken R.,
and the Stanley Neuropathology Consortium - Analysis of
RNAs from the Brains of individuals with psychiatric
Diseases (Unpublished) Stanley Neuropathology Laboratory,
Johns Hopkins School of Medicine, Baltimore MD."
BASE COUNT 140 a 61 c 115 t
ORIGIN
Query Match 62.1%; Score 368; DB 9; Length 368;
Best Local Similarity 100.0%; Pred. No. 5.3e-132;
Matches 368; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 gatagaacggaacaaatgaatgctgagttcttattaccagttgttagttgttc 94

```


Db 368 GATAGCAGCGAAGAACTCAAAATGTCGACCTTCTTATTACCACTGTTAGTATTGTTTC 309
Qy 95 tgaagaactgcttgccaagacacacattatctactgcttagaacactgcttatgttg 154
Db 308 TGGAAATGCTTGGCAAGACACATTTTATTACTGTTAGAACACTTGGCTTATGTTGTG 249
Qy 155 tgaacatacttcacacaatgltlaattatagtggtgtgaacaggaatgcacatc 214
Db 248 TGTACATATTTTCCACAAATGTTATTATTATAGTGTGTTGAACGATGCAATCTT 189
Qy 215 ttgtgtcctaaaggctgctgaagttaaaaaacaacactttcttccaatagcagtg 274
Db 188 TTGTTGCTTAAAGGTGCTGCGATTAAAAACAACTTTCTTCAATATGCGACATG 129
Qy 275 agtgaagctttttaacttaaaacacaaatglttaaatcatgtgttatcagt 334
Db 128 AGTGAATTTTAAAAACATCAAAAATGTTAAATCATGTTTATCTAGT 69
Qy 335 agttataatcgcgtatattcccatgaatgatacagaactgaacttaattcag 394
Db 68 AGTTATAATATATGCGTTTATTTCCCATGATGATGATCAACTGACATTTATTCATG 9
Qy 395 ttgtgtc 402
Db 8 TTTGCTCTC 1

RESULT 7
BE856736/c 536 bp mRNA linear EST 29-SEP-2000
LOCUS 7668a06.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:329794 3', mRNA sequence.
ACCESSION BE856736
VERSION BE856736.1 GI:10370063
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 536)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -40UP from GIBCO
High quality sequence stop: 448.
Location/Qualifiers
1. .536
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:329794"
/clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and cloneIDs: Soares NB2HP pool 1:
309384-310919, 323208-325895 Soares NB2HP pool 1:
145032-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152377 Soares NB2HP8-9W pool 1:
758280-760583, 772104-774407 Soares NBHPA pool 1:
304776-306311, 320136-322823, 326280-326653 Soares NBHOT
pool 1: 723720-726407, 739080-740999 Subtraction by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT 209 a 85 c 68 g 174 t
ORIGIN
Query Match 59.9%; Score 355; DB 10; Length 536;
Best Local Similarity 100.0%; Pred. No. 4.5e-127;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 239 aaaaaaaacacctttcttcaatagtcacatgtagtgagtttttaacttaaa 298
Db 442 AAAAAAAACACCCCTTTCTTCAATATGGCATGTAGGAGTTTAAAACTTTAAAA 383
Qy 299 acatcaaatatgttaaatcatcgtgtatcatagtaatttaataatcgctatalt 358
Db 382 ACATCAAAATTTGTTAAATATGTTGTATCTACTATTATATTATTCGGCTTATAT 323
Qy 359 tcccatgaatgatcagaactgacatttaacatgctgttcgtccatgcttctact 418
Db 332 TCCCATGAATGATCAGACATGACATTTAATTCATGTTGTGTCGCCATGCTTACT 263
Qy 419 ttaacatacttcttgcgaagaatgaaaggaatgaatagttatataagtgact 478
Db 262 TTAACATATTTCTTTGCGAATGTAAAGGTAAATGATATATATATTAAGTACT 203
Qy 479 ggcgtaaatgatgtaataatacttcttgaatgaagggcttaacagaacatgtgaac 538
Db 202 GGCTTAATATGCTTAATATATCTTATGCAATTAAGGCTTACAGAACATGTTGAAC 143
Qy 539 ttttttacttattgggaatgaagaaatgttgcacctcaacatttatgctt 593
Db 142 TTTTATTACTTTTATTTGGAATGAAGATGTTGCACTCCACATTTATTTGCTT 88

RESULT 8
BF438152/c 536 bp mRNA linear EST 30-MAR-2001
LOCUS 7667f12.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3703462 3',
DEFINITION mRNA sequence.
ACCESSION BF438152
VERSION BF438152.1 GI:11450669
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 536)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL, send email to:
info@image.lnl.gov
Seq primer: -40UP from GIBCO
High quality sequence stop: 481.
Location/Qualifiers
1. .536
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3703462"
/clone_lib="NCI_CGAP_Lu24"
/tissue_type="Carcinoid"
/lab_host="DH10B"
/note="Organ: Lung; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI_CGAP_Lu5 was prepared, and ss circles were

made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT
ORIGIN

212 a 85 c 67 g 172 t

Query Match 59.9%; Score 355; DB 10; Length 536;
Best Local Similarity 100.0%; Pred. No. 4.5e-127;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 239 aaaaaaaacaccccttcttccatgatgcatgtagagatcttcttaactttaa 298
Db 439 AAAAAAAAAACACCTTTCTTCAATATGCGATGTAGTGGAGTTTAACTTAA 380
QY 299 acatcaaaatltgtaaatcatgtgtatctagtagtataatcatcgcttata 358
Db 379 ACATCAAAATGTGTAATCATGTGTATCTAGTAGTTATTAATATGCGCTATAT 320
QY 359 tcccataatgatcgaactgacattcaatgattgtctgcgcagtccttact 418
Db 319 TCCCATGATGATGCAAGACTGACATTATTCATGTCTCGCCATCTTCTTACT 260
QY 419 ttaacatattcttgcagaatgtaaaagtaataatgattgattataatgtagt 478
Db 259 TTACATATTTCTTTTGCACAAATGTAAAGTAAATGATTAATGTTATTAAGTGTACT 200
QY 479 ggcgtgaatgacgtcaaatatacttcatgcaatgaaggcttacagaacatgtgaac 538
Db 199 GGCTGAATATGATGCAAAATATATTCTTATGCAATTAAGGCTTACAGAAATGTGAAC 140
QY 539 ttttttacttattatgggaataagaaagtgttcacccacacatttattgctt 593
Db 139 TTTTCTTACTTTTATTTGGGAATAGGAATGTTTGCACCTCCACATTTTATTTGCTT 85

RESULT 9
LOCUS AI217518/c 469 bp mRNA linear EST 17-MAR-1999
DEFINITION qh20g08.x1 Soares_NFL_T_GBC_SI Homo sapiens cDNA clone
IMAGE:1845278 3', mRNA sequence.
AI217518
AI217518.1 GI:3797333

ACCESSION
VERSION
KEYWORDS
SOURCE

ORGANISM

Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1085 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 456.

FEATURES
Location/Qualifiers

1..469

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1845278"

/clone_lib="Soares_NFL_T_GBC_SI"

/lab_host="DH10B"

/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site: 1; Not I; Site: 2; Eco RI;

Equal amounts of plasmid DNA from three normalized

BASE COUNT
ORIGIN

184 a 71 c 63 g 151 t

Query Match 58.2%; Score 345; DB 9; Length 469;
Best Local Similarity 100.0%; Pred. No. 3.5e-123;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 caacatttcttccatgatgcatgtagagatcttcttaactttaaacaacata 308
Db 428 CAACCTTTCTTCAATATGCGATGTAGTGGAGTTTAACTTAAACATCAAAA 369
QY 309 ttgttaaatcatgtgtattatctagtagttataatcatcgcttataatccccatgaa 368
Db 368 TTGTAAATCATGTGTATCTAGTAGTTATTAATGCGCTTAATTTCCCATGAA 309
QY 369 tgaatcagaactgacatttcaatgattgtctgcgcagtccttcttaacttaacatatt 428
Db 308 TGATCAGAACTGACATTTAATTCATGTTGTCTGCGCATGCTTTTACTTAACATATT 249
QY 429 tctttgcagaatgtaaagtaatgataatgattataatgaatgtagtgcgtgtaaat 488
Db 248 TCTTTTCAGATGTAAAGTAAAGTAAATGATTAATTAATTAATGATGATGCGTGAAT 189
QY 489 gatgctaaatatacttattgcaatgaaggcttacagaacatgttgaactttttact 548
Db 188 GATGCTAAATATATTCTTATGCAATTAAGGCTTACAGAAATGTGAACCTTTTACT 129
QY 549 ttattgggaataagaaatgttcacccacacatttattgctt 593
Db 128 TTTATTTGGGAATAGGAATGTTTGCACCTCCACATTTTATTTGCTT 84

RESULT 10
LOCUS AI078834/c 422 bp mRNA linear EST 10-AUG-1998
DEFINITION q246g05.x1 Soares_NhHMPU_SI Homo sapiens cDNA clone IMAGE:1678377
AI078834
AI078834.1 GI:3413141

ACCESSION
VERSION
KEYWORDS
SOURCE

ORGANISM

Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -40M13 fwd. ET from Amersham
High quality sequence stop: 361.

FEATURES
Location/Qualifiers

1..422

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1678377"

/clone_lib="Soares_NhHMPU_SI"

/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"

/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT73D-Pac
(pharmacia) with a modified polylinker; site_1: Not I;
site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2N6H, pregnant uterus
NBHPU, and fetal heart NBH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

BASE COUNT 169 a 64 c 54 g 135 t
ORIGIN

Query Match 57.5%; Score 341; DB 9; Length 422;
Best Local Similarity 100.0%; Pred. No. 1.3e-121;
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 253 cttcttccaatgacatgtagtgaggttttctaacttaaacatcaaatgtc 312
Db 422 ctttttttcaatgacatgtagtgaggttttctaacttaaacatcaaatgtc 363
QY 313 taaatcatgtgtagtctagtagttagttagttagttagttagttagttagt 372
Db 362 TAAATCATGTGTGTATCTAGTATGATTAATATCGGCTTATTTCCCATGAT 303
QY 373 cagaactgacatttaattcattgttctgcacatgcttcttaacttaacatattct 432
Db 302 CAGACACTGACATTTAATTCATGTTGTCGCCATGCTCTTTTAAACATATTTCTT 243
QY 433 ttgcagaatgtaaaaggtaattgataattgataattgataattgataattgata 492
Db 242 TTTCAGATGTAAAGGATATGATTAATTAATTAATTAATTAATTAATTAATG 183
QY 493 ctcaatattctatgtaattgaggtctacagacatgttgaacttttttaacttta 552
Db 182 CTTAAATTAATTTATGCAATTAAGGCTTACAGAACATGTTGAACATTTTTCCTTTA 123
QY 553 ttgggaataagaatgtttgcacctccacatttatgtctt 593
Db 122 TTGGGAATTAAGGAATGTTTGACCTCCACATTTTATGCTT 82

RESULT 11 410 bp mRNA linear EST 07-JUL-1998
AA954825 0095d10.s1 NCI_CGAP_K1d6 Homo sapiens cDNA clone IMAGE:1574323 3',
LOCUS mRNA sequence.
DEFINITION AA954825
ACCESSION AA954825
VERSION AA954825.1 GI:31185520
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 410)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgaps-r@mail.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmett-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/BLNL at:
www.bio.lnlnl.gov/db/rrp/image/image.html
Insert Length: 2008 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 393.
location/Qualifiers
1. 410
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1574323"
/clone_1lb="NCI_CGAP_K1d6"
/sex="mixed"
/tissue="kidney tumor"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: kidney; Vector: Bluescript SK-; Site 1:
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo df. Pooled kidney tumors. 5' adaptor sequence: 5'
GAATTCGGCACGAG 3' 3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.0 kb."

BASE COUNT 162 a 64 c 51 g 133 t
ORIGIN

Query Match 55.6%; Score 330; DB 9; Length 410;
Best Local Similarity 100.0%; Pred. No. 2.3e-117;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 264 atagcagatgtagtgaggttttctaacttaaacatcaaatgtttaaactatg 323
Db 410 ATATGGCATGTAGTGAGGATTTTAACTTTAAACATCAAAATGTTTAAATCATGTG 351
QY 324 tttatcagtagtttaataattatcgcgttatttcccatgaatgacgaactgaca 383
Db 350 TGTATCATGAGTTTAAATTAATTAATTCGGCTTATTTCCCATGAATGATGACACTGCA 291
QY 384 tttaactatgattgtctgcacatgcttcttaacttaacatattcttctgcagaatg 443
Db 290 TTTAATTCATGTTTGTCTCGCCATGCTCTTTTAAACATATTTCTTTGCGAATGT 231
QY 444 aaaggttaatgataattgattatagtgtagtgcgtgtaaatgataatgataatg 503
Db 230 AAAAGGTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 171
QY 504 ttatgcaattaaaggcttaccagaaacatgtlgaacttttttaactttatgtggaataag 563
Db 170 TTATGCAATTAAGGCTTACAGAACATGTTGAACATTTTTCCTTTATTTGGGAATAG 111
QY 564 gaatgtttgcacctccacatttatgtctt 593
Db 110 GAATGTTTGACCTCCACATTTTATGCTT 81

RESULT 12 528 bp mRNA linear EST 02-MAR-1998
AA417817 zV04h08.r1 Soares_NHMPU_S1 Homo sapiens cDNA clone IMAGE:752703 5',
LOCUS similar to gb:J03464 PROCOLLAGEN ALPHA 2(I) CHAIN PRECURSOR (HUMAN)
DEFINITION AA417817
ACCESSION AA417817
VERSION AA417817.1 GI:2079618
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 528)
AUTHORS Mammalia; Euthetia; Primates; Catarrhini; Homnidae; Homo.
TITLE Washington University School of Medicine
JOURNAL Contact: Wilson RK
COMMENT Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800

Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 393.
location/Qualifiers
1. 410
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1574323"
/clone_1lb="NCI_CGAP_K1d6"
/sex="mixed"
/tissue="kidney tumor"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: kidney; Vector: Bluescript SK-; Site 1:
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo df. Pooled kidney tumors. 5' adaptor sequence: 5'
GAATTCGGCACGAG 3' 3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.0 kb."

Fax: 314 286 1810
 Email: estewatson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 812 Std Error: 0.00
 Seq primer: -28m13 rev2 ET from Amer sham
 High quality sequence stop: 455.

FEATURES

Location/Qualifiers
 1..528
 /organism="Homo sapiens"
 /db_xref="GDB:5975341"
 /db_xref="taxon:9606"
 /clone="IMAGE:752703"
 /clone_1ib="Soares_NhMpu.S1"
 /tissue_type="Pooled human melanocyte, fetal heart, and
 pregnant uterus"
 /lab_host="DH10B"
 /note="Organ: mixed (see below); Vector: PT73D-Pac
 (Pharmacia) with a modified polylinker; Site_1: Not I;
 Site_2: Eco RI; Equal amounts of plasmid DNA from three
 normalized libraries (melanocyte 2NBHM, pregnant uterus
 NBHPU, and fetal heart NBH19M) were mixed, and ss circles
 were made in vitro. Following HAP purification, this DNA
 was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from pools of
 5,000 clones made from the same 3 libraries. The pools
 consisted of I.M.A.G.E. clones 260232-265223,
 340488-345479, and 484488-489479."
 BASE COUNT 159 a 105 c 95 g 169 t
 ORIGIN

Query Match 52.3%; Score 310; DB 9; Length 528;
 Best Local Similarity 100.0%; Pred. No. 1.1e-109;

Matches 310; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 aggaatgcaatcttctgtctcaaggtgctgcagtaaaaaaacacattctt 261
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 219 AGGATGCATCTTTGTTGCTAAGGTGCTGCAAGTTAAACCAACCTTTCTT 278
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 262 caataagcagtgagtgaggtcttctaacttaaaaaacatgaatcttaaatcat 321
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 279 CAATATGCGATGAGTGAGGATTTTAACTTTAAACATCAAAATGTTAAATCAT 338
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 322 tggatctcagtagtttaataatcgcgtcttatctcccaatgaatgacgaatga 381
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 339 TGCGTATCTAGTAGTTTAATTAATTCGCGCTTAATTTCCCGCATGATGAGAACTGA 398
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 382 catltaatcagttgtctcgcacatgctcttactttaaacaattctcttcagcaat 441
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 399 CATTTAATTCAGTGTGTCGCCATGCTCTTACTTTAAACATATTTCTTTGACGAAT 458
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 442 gtaaaagcgaatgaataatgaattatataagtgactgctgtaaatgataatata 501
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 459 GTAAGGAGTAAGTAATTAATTAATTAAGTACTGCGCTAAATGATGCTAAATATA 518
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 502 cttatgcaa 511
 ||||||||||||
 Db 519 CTTTATGCAA 528

RESULT 13
 A1499896/c 416 bp mRNA linear EST 14-APR-1999
 LOCUS tn97d12.x1 NCI_CGAP_Ut2 Homo sapiens cDNA clone IMAGE:2177495 3',
 DEFINITION mRNA sequence.
 ACCESSION A1499896
 VERSION A1499896.1 GI:4391878
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (Bases 1 to 416)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 JOURNAL Tumor Gene Index
 COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-r@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrived by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 www-bio.lnl.gov/bdrrp/image/image.html
 Insert Length: 1662 Std Error: 0.00
 Seq primer: -400P from G1bco
 High quality sequence stop: 395.

FEATURES

Location/Qualifiers
 1..416
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2177495"
 /clone_1ib="NCI_CGAP_Ut2"
 /tissue_type="moderately-differentiated endometrial
 adenocarcinoma, 3 pooled tumors"
 /lab_host="DH10B"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.85 kb. Life Technologies catalog #:
 11539-012"
 BASE COUNT 161 a 64 c 52 g 139 t
 ORIGIN

Query Match 48.9%; Score 290; DB 9; Length 416;
 Best Local Similarity 100.0%; Pred. No. 6.4e-102;

Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 aaaaatgttaaaatcatctgtgtatctcagtgatttaataatcagctatattccccc 363
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 374 AAAAATGTTAAATCACTTGTATCTAGTATTAATTAATTCGCGCTTAATTTCCCG 315
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 364 atgaatcagaactacaatttaattcagctgtgtcgcacatgctcttacttaaac 423
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 314 ATGAATATAGAACTGACATTTAATTCATGTTGTCTCCGCAATGCTTTTACTTTAAC 255
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 424 atattcttctgcagaaatgaaggaatgaataatgaattatataagtgactgctg 483
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 254 ATATTCTTTTGCAGATGTAAGGTAAGTAATTAATTAATTAAGTACTGCGCTG 195
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 484 taatgatgcttaataatcattatgcaatgaaggtcttaagaacatgctgaaactttc 543
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 194 TAAATGATGCTAAATATACCTTTATGCAATTAAGGCTTAAGAAACAATGTAATTTT 135
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 544 ttaactttatgggaataaggaatgctgacccatcattatgctt 593
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 134 TTACTTTTATGGGAATTAAGGAATTTTGCACCTCCCATTTTATTTGCTT 85

RESULT 14
 AA809783/c 404 bp mRNA linear EST 18-FEB-1998
 LOCUS nw66c04.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1251556 3',
 DEFINITION mRNA sequence.
 ACCESSION AA809783
 VERSION AA809783.1 GI:2879189
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (Bases 1 to 404)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldi, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/ldrp/image/image.html
Insert Length: 643 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 397.
Location/Qualifiers

FEATURES

source

1. 404
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:1251558"
/clone_lib="NCI-CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD+),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). CDNA synthesis was
primed with a Not I - Oligo(dT) primer
[5'-TGTACCAATCTGAAGTGGAGCGCCCTCATTTTCTTTT-3',
1. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldi."

BASE COUNT
ORIGIN

160 a 62 c 51 g 130 t 1 others

Query Match 46.0%; Score 273; DB 9; Length 404;
Best Local Similarity 99.7%; Pred. No. 2.4e-95;

Matches 323; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 270 catgtaagtgagtttttctaacttaaaacatcaaaatgttaaatcatgtgtat 329

Db 404 CATGTACTGAGCTTTTAACTTAAACATCAAAATTTGTTAAATCTTTGTAT 345

QY 330 ctgtagttataataatcggtatataatcccatgaatgatacagaactgacattat 389

Db 344 CTAGTACTTATATATATCGGCTTATATTTCCCATGATGATGAGAACTGACATTTAAT 285

QY 390 tcatgttgcctgcctgcttcttctaacttaacatatcttcttgagaagtgaagg 449

Db 284 TCATGTTTGTCCGCGCTGCTTCTTACTTAACTATTTCTTTTGAAGATGTAAGG 225

QY 450 taatgaataatgattatacaagtgactgctgtaaatgataatgataactttatgc 509

Db 224 TAATGATAATTAAGTATTAAGTACTGCTTAAATGATGCTAAATATACCTTTATGC 165

QY 510 aattaaggtctacagacatgtgaaacttttctaacttatttggaataaggaatgt 569

Db 164 AATTAAAGGCTTACAGAACTGTTGAACCTTTTACTTTTATTTGGAAATAGGAATGT 105

QY 570 ttgacactcacaatttatgtctt 593

Db 104 TTGCACCTCCACATTTATTTGCTT 81

RESULT 15

AA417676 456 bp mRNA linear EST 02-MAR-1998
LOCUS 2704408.r1 Soares_NbHMPu.S1 Homo sapiens cDNA clone IMAGE:752655 5'
DEFINITION similar to gb:U03464 PROCOLLAGEN ALPHA 2(1) CHAIN PRECURSOR (HUMAN
)'; mRNA sequence.

ACCESSION AA417676
VERSION AA417676
KEYWORDS AA417676.1 GI:2079495
SOURCE EST.

ORGANISM human.
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 456)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Insert Length: 836 Std Error: 0.00
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 383.
Location/Qualifiers

FEATURES

source

1. 456
/organism="Homo sapiens"
/db_xref="GDB:5975490"
/db_xref="taxon:9606"
/clone_lib="IMAGE:752655"
/clone_lib="Soares_NbHMPu.S1"
/tissue_type="pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT73D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbHMP, pregnant uterus
NbHMP, and fetal heart NbH19w) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

BASE COUNT
ORIGIN

131 a 101 c 81 g 143 t

Query Match 40.5%; Score 240; DB 9; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.2e-92;

Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 aggtatgaacttttggctgctaaagtgctgcaatataaaacacactttctt 261

Db 217 AGGATGCAATCTTTTGTGCTTAAAGGCTGCTGCACTTAAAAAACAACCTTTCTTT 276

QY 262 caataagcaltgtagtgagtttttctaacttaaaacatcaaaatgttaaatcat 321

Db 277 CAATATGCGCATGTAGTGAGTTTAACTTAAACATCAAAATTTGTTAAATCAT 336

QY 322 tgggtatcctgtagttataataatcggtatataatcccatgaatgatacagaatcga 381

Db 337 TGTGTATCTAGTATGTTATTAATATGCGCTTATATTTCCCATGAAATGATGAGAACTGA 396

QY 382 cattaaatcagtttgcctgcctgcttctaacttaacatatcttcttgagaagt 441

Db 397 CATTTAATTCAGTGTGTCGCGCATGCTTCTTACTTTAACAATATTCTTTGCAGAA 456

Search completed: May 22, 2002, 07:30:39
Job time: 10809 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 05:31:50 ; Search time 3328.52 Seconds

(without alignments)
3143.520 Million cell updates/sec

Title: US-09-693-205-7_COPY_1_500

Perfect score: 500
Sequence: 1 atgatttaccggaagaccat.....gcagaattgtgtatcaaa 500

Scoring table:
OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	500	100.0	12793	6	AX119931	Sequence
2	500	100.0	12793	9	AF193556	AF193556 Homo sapi
3	500	100.0	92693	9	AL157766	AL157766 Human DNA
4	499	99.8	99819	2	AC079761	AC079761 Homo sapi
5	41	8.2	174140	2	AC069017	AC069017 Mus muscu
6	35	7.0	11492	6	AX119933	AX119933 Sequence
7	35	7.0	11493	10	AF193557	AF193557 Mus muscu
8	22	4.4	39760	10	U29187	U29187 Mus muscu
9	21	4.2	10334	1	AE006329	AE006329 Lactococ
10	21	4.2	47573	3	AF030694	AF030694 Plasmodiu
11	21	4.2	130192	9	AL157832	AL157832 Human DNA
12	21	4.2	163721	2	AC099718	AC099718 Mus muscu
13	21	4.2	169234	9	AC018633	AC018633 Homo sapi
14	20	4.0	1295	3	AF034998	AF034998 Hyphantri
15	20	4.0	57205	2	AC107995	AC107995 Homo sapi
16	20	4.0	68654	2	AC105967	AC105967 Mus muscu
17	20	4.0	77836	2	AC022573	AC022573 Homo sapi
18	20	4.0	111714	2	AL359432	AL359432 Homo sapi
19	20	4.0	134977	9	AC002347	AC002347 Homo sapi
20	20	4.0	143914	2	AC022711	AC022711 Homo sapi
21	20	4.0	151992	2	AC009939	AC009939 Homo sapi
22	20	4.0	152529	2	AC008863	AC008863 Homo sapi
23	20	4.0	153665	2	AL359743	AL359743 Homo sapi
24	20	4.0	157864	2	AC034303	AC034303 Homo sapi
25	20	4.0	159865	2	AC011568	AC011568 Homo sapi
26	20	4.0	163989	2	AP002416	AP002416 Homo sapi
27	20	4.0	167619	9	AC012148	AC012148 Homo sapi
28	20	4.0	171076	9	AC090231	AC090231 Homo sapi
29	20	4.0	177510	2	AC091871	AC091871 Homo sapi
30	20	4.0	177951	2	AC074383	AC074383 Homo sapi
31	20	4.0	177969	9	AC068107	AC068107 Homo sapi
32	20	4.0	178255	9	AC019197	AC019197 Homo sapi
33	20	4.0	180915	9	HS495L19	AL442637 Homo sapi
34	20	4.0	187449	9	AC016450	AC016450 Homo sapi
35	20	4.0	188643	9	AC092422	AC092422 Homo sapi
36	20	4.0	189006	2	AC103467	AC103467 Rattus no
37	20	4.0	189157	2	AC011129	AC011129 Homo sapi
38	20	4.0	203793	2	AC016263	AC016263 Homo sapi
39	20	4.0	213296	2	AC067784	AC067784 Homo sapi
40	20	4.0	217250	2	AC018595	AC018595 Homo sapi
41	19	3.8	721	6	A43543	A43543 Sequence 11
42	19	3.8	724	11	G57985	G57985 SHGC-103784
43	19	3.8	746	8	LETRAS14	X51904 Lycopersico
44	19	3.8	1075	8	AF068875	AF068875 Glycine m
45	19	3.8	1076	8	AF068874	AF068874 Glycine c

ALIGNMENTS

RESULT 1
AX119931
LOCUS AX119931 12793 bp DNA linear PAR 11-MAY-2001
DEFINITION Sequence 1 from Patent WO0129266.
ACCESSION AX119931
VERSION AX119931.1 GI:14036678
KEYWORDS
SOURCE
ORGANISM human.
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 (bases 1 to 12793)
Hudson,T.J., Engert,U. and Richter,A.
Identification of arascs mutations and methods of use therefor
Patent: WO 0129266-A 1 26-APR-2001;
JOURNAL MCGILL UNIVERSITY (CA) 1 26-APR-2001;
location/Qualifiers

FEATURES
source
1..12793
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT	4163 a	2256 c	2487 g	3887 t
------------	--------	--------	--------	--------

BASE COUNT 4163 a 2256 c 2487 g 3887 t
ORIGIN

[illegible]

repeat_region	64695..64713	/note="TIGER1 repeat: matches 29..46 of consensus"
repeat_region	65088..65395	/note="L1PB2 repeat: matches 5405..5733 of consensus"
repeat_region	65396..65569	/note="AluV repeat: matches 136..309 of consensus"
repeat_region	65571..65640	/note="L1PB2 repeat: matches 5728..5791 of consensus"
repeat_region	65696..65717	/note="11 copies 2 mer ta 100% conserved"
repeat_region	65725..66096	/note="L1PB2 repeat: matches 5789..6155 of consensus"
repeat_region	66371..66410	/note="10 copies 4 mer tgt 82% conserved"
repeat_region	67586..67886	/note="AluB repeat: matches 1..299 of consensus"
repeat_region	69748..69930	/note="MIR repeat: matches 6..248 of consensus"
repeat_region	70957..71267	/note="AluV repeat: matches 1..311 of consensus"
repeat_region	71279..71413	/note="MER1B repeat: matches 548..680 of consensus"
repeat_region	71411..71737	/note="MER1A repeat: matches 47..485 of consensus"
repeat_region	71780..72075	/note="AluX repeat: matches 1..295 of consensus"
repeat_region	72145..72256	/note="MER11-internal repeat: matches 42..175 of consensus"
repeat_region	72454..72865	/note="MER11-internal repeat: matches 332..739 of consensus"
repeat_region	72873..73249	/note="MER11-internal repeat: matches 883..1261 of consensus"
Query Match	100.0%;	Score 500; DB 9; Length 92693;
Best Local Similarity	100.0%;	Pred. No. 1,4e-256;
Matches 500; Conservative	0; Mismatches	0; Indels 0; Gaps 0;
OY 1	atgatttacagaagaacccatgtacacagctcgaagcttaacccaagaattgcacg	60
Db 18909	ATGATTTCAGAGAACCATGTACTACCTGACCTGACGCTTTAATCCAGACATTTGCACG	18850
OY 61	tctltacagaagaatgaatgatacatcttcgctcgtgcagagaattgattgtcaatgta	120
Db 18849	TCTTATCAAGGAGTAATGATACATCTCTGCGCTGGCAGAGAATGATTGTTCAATGTA	18790
OY 121	tcacattgaatgaagaacagaatcacccatctgtttcatatgcttaagatggtttggaataa	180
Db 18789	TCCATTATGATGAAAAACAAGAAATCACCCATCTGTTTCATGAGCTTAAGATGGTTTGGAAAA	18730
OY 181	tcttatatacatttttaagagatttgaactttatattgaatgaatgccaactatccccag	240
Db 18729	TCTTATATACATTTTTCAGAGATTTCATTTATTTGATAGAATGCCACTTATCCCAAG	18670
OY 241	aactactagaagaagtcagacatgctgtgaaacttaagacttaagatcaggaattcatcggt	300
Db 18669	AACCTACTAGAGGAGGTGACACTGTGTGGAATCTATTAGACTCAGGATTCATCGTT	18610
OY 301	agtcattttaagacgatgaatctgaaagcagcttcacagaattttaagacagacattgtaca	360
Db 18609	AGTCATTTTAGACGATGTGATCTGAAGCACACCTTCCAGAAATTTTTCAGACGACATTGTACA	18550
OY 361	aaacttggagaggttgccttaaaaaattgatagatcctataaaaaatccgctattaa	420
Db 18549	AAAACCTTGAGGGTGTGCTTAAAAAATTTGATGCACTTATACAACTCCGCTTATTAA	18490
OY 421	aaaatatatcatcaccattaccagaatgctgttttgacataatggagaagaatgacat	480
Db 18489	AAAAATATTTTCATTCACCATTAACAAGTCGTTTTCAGATTAATGAGAAAGATGCCATT	18430
OY 481	gcagaaatgtgtaatacaaa 500	


```

misc_feature      10026..11426
                    /note="assembly_name:Contig42"
misc_feature      11527..11266
                    /note="assembly_name:Contig44"
misc_feature      13367..114794
                    /note="assembly_name:Contig47"
misc_feature      14895..16054
                    /note="assembly_name:Contig48"
misc_feature      16155..17395
                    /note="assembly_name:Contig49"
misc_feature      17496..19287
                    /note="assembly_name:Contig50"
misc_feature      19388..21294
                    /note="assembly_name:Contig51"
misc_feature      21395..22944
                    /note="assembly_name:Contig52"
misc_feature      23045..24421
                    /note="assembly_name:Contig53"
misc_feature      24522..25870
                    /note="assembly_name:Contig54"
misc_feature      25971..27230
                    /note="assembly_name:Contig55"
misc_feature      27331..28778
                    /note="assembly_name:Contig57"
misc_feature      28879..30893
                    /note="assembly_name:Contig58"
misc_feature      30994..32460
                    /note="assembly_name:Contig59"
misc_feature      32561..33984
                    /note="assembly_name:Contig60"
misc_feature      34085..35285
                    /note="assembly_name:Contig61"
misc_feature      35386..37184
                    /note="assembly_name:Contig62"
misc_feature      37285..39172
                    /note="assembly_name:Contig63"
misc_feature      39273..40874
                    /note="assembly_name:Contig64"
misc_feature      40975..42893
                    /note="assembly_name:Contig65"
misc_feature      42994..44384
                    /note="assembly_name:Contig66"
misc_feature      44485..45999
                    /note="assembly_name:Contig67"
misc_feature      46100..48669
                    /note="assembly_name:Contig68"
misc_feature      48770..50798
                    /note="assembly_name:Contig69"
misc_feature      50899..52809
                    /note="assembly_name:Contig70"
misc_feature      52910..55127
                    /note="assembly_name:Contig71"
misc_feature      55228..58087
                    /note="assembly_name:Contig72"
misc_feature      58188..61004
                    /note="assembly_name:Contig73"
misc_feature      61105..64185
                    /note="assembly_name:Contig74"
misc_feature      64286..67105
                    /note="assembly_name:Contig74"

```

Query Match 99.8%; Score 499; DB 2; Length 99819;
 Best Local Similarity 100.0%; Pred. No. 4,8e-256;
 Matches 499; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 atgatttaagaagaacatgtaactcagctgcagctcttaaatcagaacgatttgcaag 60
    |||||||
DB 48099 ATGATTTACGAGAACCACTACTCAGCTGACAGCTCTTAATCAAGAACGATTGCAAG 48158
    |||||||
QY 61 tcttatacaagaagtaatgatactctgctgcagagaattgatttcaatgta 120
    |||||||
DB 48159 TCTTATCAAGAGATATGATATCAATTCCTGCGCGAGAGAAATGATTTCTCAATGCTA 48218
    |||||||
QY 121 tccatttgatgaataacgaatacccatctgttcatggttaagaatggtttgaaaaa 180
    |||||||

```

```

DB 48219 TCCATTGATGAAACAGAAATCAACCATCTGTTTCATGCGCTTAAGATGCTTGGAAAAA 48278
    |||||||
QY 181 tctttatacatcttttcaagaagatttgacttatttgagaatgacacttcccccag 240
    |||||||
DB 48279 TCTTTATATACATTTTTCAGAGAGATTTGACTTTATTTGATGAATGCCACTTATCCCCAG 48338
    |||||||
QY 241 aactatactagagaagtgtaagacatgctgtaacatcattagactcagaattccalcgt 300
    |||||||
DB 48339 AACTATATCTAGAGAGAGTGTAGACATGTGTGAACTCATTTAGACTCAGATTCATCTGTT 48398
    |||||||
QY 301 agtcattttagaagatgtaatcgtgaagacacagcttccagaatttttagcagacatgaca 360
    |||||||
DB 48399 AGTCATTTTAGACGATGAATCTGAAGCAGCTTCCAGATTTTATAGACAGACTTGTACA 48458
    |||||||
QY 361 aaactctggaggttcttctcttaaaaaataagatgatactcatacaacatcggcttataa 420
    |||||||
DB 48459 AAAACCTTGAGAGGTTGCTTCTTAAAAAATGATGATCTATACAACTCCGCTTTTAA 48518
    |||||||
QY 421 aaatatattcatcacattaccacgaagtgtgttttgcagataatgagaagatgacatt 480
    |||||||
DB 48519 AAAATATATTCATTCACCATTAACCAAGTCTGTGTTTGCAATATATGAGAAAGATGCCATT 48578
    |||||||
QY 481 gcagaatgtgttaatcaa 499
    |||||||
DB 48579 GCAGAAATGTGTATCAAA 48597
    |||||||

```

```

RESULT 5
AC069017 174140 bp DNA linear HTG 12-SEP-2001
LOCUS Mus musculus clone MGS3-342116, WORKING DRAFT SEQUENCE, 5 unordered
DEFINITION
AC069017
VERSION AC069017.20 GI:15559167
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE 1 (bases 1 to 174140)
AUTHORS Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
Dederich,D., Thomas,S., Okwuonu,G., Carllock,C., Garner,T.,
Addison,S., Pace,A., Williams,G., Bonin,D., Brooks,A., Brown,J.,
Buhay,C., Bunnell,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
Fernandez,C., Ferraguto,D., Fortcum-Tansey,J., Gill,R.,
Gorrell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hognes,M.,
Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
Kovar,C., Liu,J., Liu,W., Louisedge,H., Lozado,R.J., Martin,R.,
Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Ogih,M., Parish,B.,
Perez,L., Reiter,D., Say,J., Shen,H., Vasequez,L., Wallington,S.,
Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstock,G.,
Worley,K. and Gibbs,R.

```

```

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 174140)
AUTHORS Worley,K.C.
JOURNAL Direct Submission

```

Submitted (17-MAY-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

```

COMMENT ----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: MGAO
Center clone name: MGS3-342116

```


SVNDCITLYQENITNACRYLHEAVLQNMAMATITIEKIPCEFIILVENYVESEKVS
FHLNFEARAYLYQLPKNKYKNNRELESEVQSQSFVEFPAVLVEIDDERKQJTE
ENFOLCRITSEIGNLLREKREOECEKNGVGLIPDLMLLPALSLCYNCPMTKY
KSDJYKVCADIPREYAVAKGLAIKPRKHLKELEVASICITGALTEGOKELTSRKS
ILNAYPSEKEMKELLQNMDDAKATEICFPPDROPHVRITDDKAPLOGLACVYN
NOPFTEDDYRGIONIGKCTEGNPKCTGHGIGFNSVYITTCPPSTISNDLIGTIDP
HARVAPGATVSPGRNFRDLADPFOFSDVLDLYGNFKLDNDCTMEFPPLRNAMA
QVSELSVSPSSDRMVOQLDLRSODAEIIMLNHEKISIEIDENCTMFPPLRNAMA
GKITDDGRKLRKQFHASVYDSYTKRQKLDIVQOQITTMDEDSGNTLTMICRS
GFSMEKVSYSIAHKNDITLPPRGGAACITTHYKCPKHPAFGLPISLEGLPEFH
VNGHFDLQSRRLNRDDMGVGRSDMNSLTALAPAYVLLIOLKRRYPPGSPPT
LSVLOQTPHYKDTLKKFLSFPVNRDLQDPLCLVALNLSCHIEDKRLIPVRA
PNIDSDLSHSAVITWINSTNKTREPDNLQDELQALNKADVNTTRKVAENVY
RLKHLLEIGFNLVNCDETANLHOLVDADIPVSPADVRSFLMTPSSPDNCHI
GLPCRLQOTNKLFEHSLKLDVYCGKDESEFEVEGLPLITLTDVQIOPDGRPK
FLTTYHELIPSRKDLFMTLYLKSSVLDLNCVAKVDFIISRPALDLSYLPREYKKN
CAKMDNPAESWMLKNAHMFISESVSYVDQDEPRAPVYIDILKDALILGTRKY
STISQLYPBGDVLIPLSLHIAVFPNAGSDKYFHALKAGCIOLALNKICSDALVP
LISCHTANIDSPASILKAVYVOTSTFTEKLMENDFALLMYFNCNLSHMSODDI
KILKSLPCYKSSISGRMSIAKFGTCVYLKSPSAEVEKQSSGSAFLSEKYLKEL
YVLGCVPDQDLEVYLKLLPKITENLSYAKLEHLIYLKRLASIEPESEIKOLEP
LESLLIHDANNRLKOAKHFYDRTVRFVEMLPBKFLIPKFEFKLEOYIKRNOAF
MISWYEFLLNIGLKYALSOOOLQFRAKETSANENMSEKELQSVYDILHIIHER
MDLSGNFLKEILSIPFLCPEPAPAEIIRFHQJOENVGTLPLKFNAGVNPKEKQC
DVLQDLMTSCPLPEKAPPULSTKEQESDLAEQOLEQVLMNLVNDLPLKVINNC
RNICHTITLDEMNKTRAKVLRISITIEFLSAERREFFOGLVYAVFVEDGWLKPEE
VYINLEYEADFPYLYKLPLELGTGHQLEHGTEDISTKOYEVLSIFKSSSEKO
LDPNEMRYKRVYSGFLKSDNSVYKRDNLNARLALYLPDSODKLVKSSILVDD
APHYKSRIOGNIGVOMLVDSOCYLGKDGHPKTLIMFEPKRLKESLISIEODE
ETPKVQORALCSLOGRLOLLSSEQFITGLIRIKHENDNAPLANEERAIKRLKLR
EGLKVSCEFKLOTLTKVGFNPISRSRSTFALRFGNAVILLIYOHSDSDINLL
ALAMTKSATDMLISDTSYLIAMLCNDIYRISSEKLDAGVYDSSEPKLELPMFGT
PIPAIHLYLMDPMNVFPGYVGLVDAEGDLYGSQPTYVAIYQVEEREDAD
NPSFGIKYQIODIGSEKIVSSIDLKFSRDESSQNRDSAPTPPTSPTELEPBLR
SLPPLFSGESHKSPSTKHSRKLKVNALPRLKEVTSVQAKMLPSEKKIIR
LYLKHPDNPNPNHDIANEVFKHLQLEIRLKOAFLDONAPRASRITSTASRQUS
DKYSFORFYSWNOQETSHKSRQOQKCRKCPSSAGQTYSPFKSVGNPEA
RMLROARANFSAARNDLKNNANENWCFKYSTKLALIAAYAVAGKSDKVKPAL
AOKIEEYSQOLEGLTNDVHTLEAYGVDSLKTRYPDLLPPOQIPNDRTSEVAMRVMEC
TACITIKLENFTOOKY"

BASE COUNT 3599 a 2281 c 2387 g 3226 t
ORIGIN

Query Match 7.0%; Score 35; DB 10; Length 11493;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 203 gattgacttattatgagatgcactatccc 237
Db 127 GATTGACTTATTGATGAGATGCACCTATCCC 161

RESULT 8
LOCUS U29187 39760 bp DNA linear ROD 18-JAN-2000
DEFINITION Mus musculus long incubation prion protein (Prnpb) and prion-like
ACCESSION U29187
KEYWORDS U29187.1 GI:5281065
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 39760)
Lee,I.Y., Westaway,D., Smit,A.F.A., Wang,K., Seto,J., Chen,L.,
Acharya,C., Ankener,M., Baskin,D., Cooper,C., Yao,H., Prusiner,S.B.
and Hood,L.E.
TITLE Complete genomic sequence and analysis of the prion protein gene
JOURNAL region from three mammalian species
Genome Res. 8 (10), 1022-1037 (1998)
MEDLINE 99018115

PUBMED 9799790
REFERENCE 2 (bases 1 to 39760)
AUTHORS Moore,R.C., Lee,I.Y., Silverman,G.L., Harrison,P.M., Strome,R.,
Heinrich,C., Karunaratne,A., Pasternak,S.H., Chishti,M.A.,
Liang,Y., Mastrangelo,P., Wang,K., Smit,A.F.A., Katamine,S.,
Carlson,G., Cohen,F.E., Prusiner,S.B., Melton,D.W., Tremblay,P.,
Hood,L.E. and Westaway,D.
Ataxia in prion protein (Prp)-deficient mice is associated with
upregulation of the novel Prp-like protein dopple
J. Mol. Biol. 292 (4), 797-817 (1999)
JOURNAL 99457485
MEDLINE 10525406
PUBMED 10525406
REFERENCE 3 (bases 1 to 39760)
AUTHORS Lee,I.Y.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-1995) Department of Molecular Biotechnology,
University of Washington, Box 352145, Seattle, Washington
98195-2145, USA
4 (bases 1 to 39760)
Lee,I.Y.
TITLE Direct Submission
JOURNAL Submitted (31-AUG-1999) Department of Molecular Biotechnology,
University of Washington, Box 352145, Seattle, Washington
98195-2145, USA
REMARK COMMENT
REMARK Sequence update by submitter
COMMENT Interspersed Repeats were identified with RepeatMasker (available
from <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).
FEATURES
source
1..39760
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="2"
/clone="cos6.1/LnJ"
/note="strain=1/LnJ"
complement(1..167)
/rpt_family="ORR1B"
complement(1258..1677)
/rpt_family="MER67C"
1751..1863
/rpt_family="B4"
1857..1925
/rpt_family="(CA)n"
1874..1925
/rpt_family="(AC)x25
2184..2500
/rpt_family="ORR1A3"
complement(2866..3091)
/rpt_family="B3"
4902..4940
/rpt_family="(CA)x19
complement(5239..5288)
/rpt_family="B4"
complement(5241..5326)
/rpt_family="RSTINE1"
complement(5371..5422)
/rpt_family="PB1D10"
5572..5712
/rpt_family="B1_MM"
5572..5561
/rpt_family="PB1D7"
6205..6251
/gene="Prnpb"
/number=1
join(6205..6251,8442..8539,19667..21675)
/gene="Prnpb"
/product="long incubation prion protein"
6205..21675
/gene="Prnpb"
repeat_region complement(7213..7437)
/rpt_family="B3"
repeat_region complement(7637..7756)
/rpt_family="B1-F"
repeat_region complement(7663..7756)

exon	/rpt_family="PB1D7" 8442. .8539 /gene="Prnpb" /number=2
repeat_region	complement(8894. .8945) /rpt_family="PB1D10"
repeat_region	complement(10249. .10341) /rpt_family="ID_RN"
repeat_region	complement(10269. .10341) /rpt_family="ID4"
repeat_region	complement(10598. .10808) /rpt_family="B2"
repeat_region	11198. .11343 /rpt_family="B1_MM"
repeat_region	complement(11573. .11651) /rpt_family="ID_RN"
repeat_region	complement(11592. .11651) /rpt_family="ID5"
repeat_region	complement(12621. .13162) /rpt_family="LIM44"
repeat_region	complement(13163. .13262) /rpt_family="PB1D10"
repeat_region	13809. .14171 /rpt_family="OR1D"
repeat_region	complement(14242. .14376) /rpt_family="B1_MM"
repeat_region	complement(15937. .15985) /rpt_family="ID6"
repeat_region	16797. .16879 /rpt_family="LTR10A"
repeat_region	complement(17191. .17372) /rpt_family="LINE2"
repeat_region	17438. .17462 /rpt_unit=(GATT)x6
repeat_region	complement(17460. .17526) /rpt_family="ID4"
repeat_region	complement(17960. .18149) /rpt_family="L1_MM"
repeat_region	18517. .18612 /rpt_family="PB1D7"
exon	19667. .21675 /gene="Prnpb" /number=3
CDS	19677. .20441 /gene="Prnpb" /codon_start=1 /product="long incubation prion protein" /translation_id="AAD41440.1" /db_xref="GI:5281066" /translation="MANLGWYLALEFVMTDVGCKRKPKPGGMNTGGSRYPPGSGS GGRNPPGGGOWPHGGGOWGPHGGSGWGPBGHGGGOGGGGCTPNOMKRS KPKTRFKIVAGAAAGAYVGLGCGMLGSAASRPIHFNGNEMERYTRENNYRIPNOV YRPVFDKISNOGNEVITIKOHTVVTTKENFETDYKMMERVEQMLCVTORQ RESQAYVGGSSNVFLFSSPPVILLISFLFLIVG"
repeat_region	complement(21935. .22120) /rpt_family="B3"
repeat_region	complement(22278. .22387) /rpt_family="B1-F"
repeat_region	complement(22419. .22548) /rpt_family="B1-F"
repeat_region	23599. .23872 /rpt_family="B4"
repeat_region	complement(24414. .24565) /rpt_family="MER5A"
exon	25842. .26004 /note="Intergene exon 1" complement(27124. .27269) /rpt_family="B1_MM"
repeat_region	complement(27289. .27350) /rpt_family="B1_MM"
repeat_region	27385. .27520 /rpt_family="B1_MM"
repeat_region	27648. .27833 /rpt_family="B1_MM"

FEATURES	LOCATION/Qualifiers
repeat_region	/rpt_family="MLT2D" 27819..27910
repeat_region	/rpt_family="(GCA)n" 27828..27885
repeat_region	/rpt_unit="(GCA)x19 28003..28120
repeat_region	/rpt_family="MLT2CB" complement(28530..28729)
gene	/rpt_family="MLT1B" 29572..29672
exon	/note="Intergene exon 2" 30254..30345
repeat_region	/rpt_family="LIME3" 30341..30573
gene	/rpt_family="LIME3A" 34086..39315
exon	/gene="Prnd" /note="prion-like gene" 34086..34124
exon	/gene="Prnd" /number=1 join(34086..34124,36205..37472,38013..39315)
mrna	/gene="Prnd" /product="prion-like protein" 35130..35341
repeat_region	/rpt_family="URR1A" 36205..36799
exon	/gene="Prnd" /note="exon 2a: short form of exon 2" 36205..37472
exon	/gene="Prnd" /note="exon 2b: long form of exon 2" 36212..36751
CDS	/gene="Prnd" /codon_start=1 /product="prion-like protein"
Query Match	4.4%; Score 22; DB 10; Length 39760;
Best Local Similarity	100.0%; Pred. No. 14;
Matches	22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Ox	26 cagctgcagctctcaatccag 47
Db	25620 CAGTGCAGCTCTCAATCCAG 25641
RESULT	9
LOCUS	AE006329 10334 bp DNA linear BCT 14-MAY-2001
DEFINITION	Lactococcus lactis subsp. lactis IL1403 section 91 of 218 of the complete genome.
ACCESSION	AE006329 AE005176
VERSION	AE006329.1 GI:12723880
KEYWORDS	
SOURCE	
ORGANISM	Lactococcus lactis subsp. lactis. Lactococcus lactis subsp. lactis Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Lactococcus. 1 (bases 1 to 10334) Bojotin,A., Winker,P., Mauger,S., Jallion,O., Malarme,K., Weissenbach,J., Ehrlich,S.D. and Sorokin,A. The Complete Genome Sequence of the Lactic Acid Bacterium Lactococcus lactis ssp. lactis IL1403 Genome Res. 11 (5), 731-753 (2001)
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	
PUBMED	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	

source
1. .10334
/organism="Lactococcus lactis subsp. lactis"
/strain="LI1403"
/sub_species="lactis"
/db_xref="taxon:1360"
71. .1129
/gene="yJha"
/note="LI69897"
71. .1129
/gene="yJha"
/note="EVIDENCE BY GENEMARK BIO16.01 Unknown"
/codon_start=1
/transl_table=11
/product="UNKNOWN PROTEIN"
/protein_id="AAK05042.1"
/db_xref="GI:12723881"
/translation="MKNRKLKIGKILVSIILFPALITGSYVIIVYVQVIRLPDMKOT
IKNPFOVELQOTYKMTFNIGYSYSPSEFMGSDGVRAVSKSAVSAIOEDIR
LINKRPDGNIOEIDMOGRSQOVNEPOMVRAGLPDYTSVLQNDYSAYLEPYTPK
ICKASGIMTVSKYEIESSTRYKLPETNENKFFDLRAFSVILPIKNSDKFVLEN
THLSAFTDOKIOKQOLFLFDEMKKYVSKGDYVIGADYNHNLAKKAPDELTKMEF
PTENLTGMRVAVPTNAPTVRSIDVAYDKKNPKNTEGIIIDFLVSNINDIKQITIDN
OFKSSDHQPVWVDFSEK"
complement(1155. .2096)
/gene="yJhb"
/note="LI171002"
complement(1155. .2096)
/gene="yJhb"
/note="EVIDENCE BY HOMOLOGOY BIO15.01 HYPOTHETICAL.
General. There are three paralogs in L.lactis. 238
identical to putative 1,4-dihydroxy-2-naphthoate
octaprenyltransferase of B.subtilis."
/codon_start=1
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="AAK05043.1"
/db_xref="GI:12723882"
/translation="MTLKTFFNFTRIQIPLPALISPIACCLFAFWYGSFHEMPITLF
FIGLASINLFVAMNNIMDYOKALDEPEYKNNIVKSNPKLALNICALLAIDVV
GLIVLEITMLILPGLICELIAIETVGPAPSPRPLGELLAGEPGEFLAYVI
NSYDKGFPIINDRNMTWMTWDFKILPLILVGLMCFVQNRINMSDNICDEQIR
NERTYLPYTLKPLSLKVTGTLFAGLCLILAIAGILPIWSLMSITIALKVTPIK
KPRAKOSKSETHLQIDNMLFNASLALTLGLITFK"
2250. .2513
/gene="yJhc"
/note="LI172073"
2250. .2513
/gene="yJhc"
/note="EVIDENCE BY HOMOLOGOY BIO15.02 HYPOTHETICAL.
Conserved. 488 identical to Mj1558 of Methanococcus
jannaschii. Probably F0F1-type ATP synthase alpha subunit,
evidenced by cognitor."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK05044.1"
/db_xref="GI:12723883"
/translation="MRAYVTVGADKIGIVAGVTATLAELENIIFISQTLMSGFTM
MMVVEIQNSDFSAFQALSRKEBALGVNIHVQNEALFNAMHKL"
2630. .3967
/gene="yJhd"
/note="LI172471"
2630. .3967
/gene="yJhd"
/note="EVIDENCE BY HOMOLOGOY BIO15.01 HYPOTHETICAL.
Conserved. 598 identical to Mj1665 protein of
Methanococcus jannaschii."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK05045.1"
/db_xref="GI:12723884"
/translation="MDIONIKETIAMIEBQNFDIRITMGISILDCIDADIDAKAEKI

gene
YOKIVKAGKLVENGEIGHELGIKIVNRKVSVPPIAIGATTAADYTPPLALAMDRA
AKEIGIDFIGYSALVQKGYOKGDEILISMFKALATERRVCAVAVGSGTKGIMNTA
VRMGETIKMSGDKMLNKLAVFANVAEDNEPFGAGHVEGAPTIIINVGSIGTA
VKRALEKVGESPDILAEITKTKAFKITITGOLVQGMASERINVEGIDYLSAPPA
IGDSVARVLEBMLETVGTHTTALAMANDAKKGVAAERVGLSAPFVPSDE
GMTAAVNSGALNLEKLEAMTCCSVGLDMITAIPEETPASTIAMADEAIGVINOKT
TAVRIIPMGEQIEFGILFGVAPVRVKNKSSADFIARGQIIPAIHSEFN"
4067. .4546
/gene="yJhe"
/note="LI173848"
4067. .4546
/gene="yJhe"
/note="EVIDENCE BY GENEMARK BIO16.01 Unknown"
/codon_start=1
/transl_table=11
/product="UNKNOWN PROTEIN"
/protein_id="AAK05046.1"
/db_xref="GI:12723885"
/translation="MAYKVFNKDVTETVGLTSPVADSLAGLRANEARVFNKYFEYL
TYPAHKADEVALFEKILLEERLAKFSARILAEVATENDELYWPEFYVEDGVALNLLY
EKVGQGRKPRKRAVGIKLSVGMETPRELEBKFPKFAKROSKLAGEIRGSYFTIKQEWL"
4614. .5309
/gene="yJhf"
/note="LI174407"
4614. .5309
/gene="yJhf"
/note="EVIDENCE BY HOMOLOGOY BIO06.09 Glycolysis"
/codon_start=1
/transl_table=11
/product="phosphoglycerate mutase (EC5.4.2.1)"
/protein_id="AAK05047.1"
/db_xref="GI:12723886"
/translation="WVNIYLVIRHGTNFTNTIGRAQGSMDPTPLTAGEOGIIEGLGFL
EKIGKPRDRAVSSDSGRTIOTMGFIILANSNEGIPRYDKRIRPWCMSGSDGDELF
DGYLPRIFTETGHNHTRDERPPDEVINAKLFEVDYDAGMAPHELSRINDGFTDIKE
SEAKKANIVIVSHGWTIGTFVKLAHPELPRPHGLDNGSVTHLSEFDGKFTTGKVDG
SYREIGSEKLRK"
5306. .6052
/gene="dacB"
/note="LI175147"
5306. .6052
/gene="dacB"
/note="EVIDENCE BY HOMOLOGOY BIO03.02 Murein sacculus and
peptidoglycan. in E.faecalis corresponding enzyme mediates
vancomycin resistance"
/codon_start=1
/transl_table=11
/product="D-alanyl-D-alanine carboxypeptidase (EC
3.4.16.4)"
/protein_id="AAK05048.1"
/db_xref="GI:12723887"
/translation="MKRRITFSLIVLVILIGALGPIILITNDKSNSTNSAKKTSS
SKEVAKKRLDPAKQSDMDLVYLSKSPDENKPTLATYNGIOVDRIASAVQOPLAA
AQGISPAEHLISGYSVSYQEGLYOSIVQAKENGGQTVNSGSPISSEDAIKNVOTIS
OPRRSSEHQGTGLAIDMSNIDLSIESQYAKOVAIAIRKGFVLRPFGKRAETGIGED
WHPRYVGLENAKYMTENHLTLEBYKLALPA"
6066. .6470
/gene="yJih"
/note="LI175880"
6066. .6470
/gene="yJih"
/note="EVIDENCE BY GENEMARK BIO16.01 Unknown"
/codon_start=1
/transl_table=11
/product="UNKNOWN PROTEIN"
/protein_id="AAK05049.1"
/db_xref="GI:12723888"
/translation="MLNRRRLLYVVALFIYSSEMIKVLVHNHFPQSOFNNLSTN
IKTITKISSITSMOSIYILITFALLILKFMFTFILLEITISALLIDYALIFI
QHEVYKGSFLVIANVYITVANNIQTSDSNLL"
6691. .7734
/gene="hrcA"
/note="L0212"


```

/evidence=experimental
/product="C3"
/protein_id="AAC47847.1"
/db_xref="GI:2642505"
/translation="MSDVSIDKEESYSKDEYISEKKESLASHKNDREENEVE
EKVEKVDKHLLEKLEKMKSSVKSDNSMKNKEIYOLAKMSDEKREELKRLKE
KRRKREIYDGESEDEGHDKRYFNKNLNLCESKNDEGHDKRYFKKKK"
complement(9741..10358)
/gene="C2"
complement(9741..10358)
/gene="C2"
complement(9741..10358)
/gene="C2"
/codon_start=1
/evidence=not_experimental
/product="C2"
/protein_id="AAC47846.1"
/db_xref="GI:2642504"
/translation="MENKMKINAFYKPVVEIKRYVHKGDTHLYKLYEDPKNKDK
KYKMKNEYTKLKNHILNMINKKKEGCVANNKIIVHFLFNEDVTPKIEIVY
IKRLEKNIENVSHLEKNVNDYTVKPVVPIEVPILKFRDHFKIIPIRKKIIPVI
KYTDNIYVDCCEKRYIYENVVIPPVPCDIPLEOKKXIDKVPPI"
11113..11322
/rpt_family="microsatellite 1H6"
complement(11754..12377)
/gene="Cg8"
complement(11754..12377)
/gene="Cg8"
/codon_start=1
/evidence=experimental
/product="Cg8"
/protein_id="AAC47845.1"
/db_xref="GI:2642503"
/translation="WKIFRRRYVASQPLGINIPDPPEYLPNGEPKTYRLNQRYN
HKYDERMTLRPAGVHTYVGEFAKDHLYPLAFLIYPLFAQLSPAF
HHISYGIKPRPTEVMEVANKREBAENNTPTRLDRRKRGRGHWLLDEPLSPHP
CYLHMSKYHHDELLKMEBEAEKAESEDEEEDDITSEDDN"
12462..12639
/rpt_family="microsatellite 3E7"
/rpt_unit=tata
16000..18567
/gene="Cg4"
16000..18567
/gene="Cg4"
/codon_start=1
/evidence=experimental
/product="Cg4"
/protein_id="AAC47838.1"
/db_xref="GI:2642496"
/translation="MSVLCIDIGDNDNSVATINKAINVVRDISRLAPLYGFTFK
ERLIGDSALSKLSNKTCKRNKILGIGDVKDIDIEHAYGDLICEYVLYGE
VEYKNEKVVSAVRLSALSLIKMAERYIGCEKEIYLSPPPTNOKBCLLAT
KIINANVRLIISDNTAVADGYMRKPERKEDGSLVFNIGYANTCVARFENK
CEILDADSNLGGRLDNEILKIYINIVNNYKMPLYKNNPDELCDPGGSLNLF
VSTASDOONGNNRYRIKLOEVAIKTKRVLSANNASIHVECIYDLDGGSINER
FEELCSNPLUTKHLIDTALCISKYNIODIHSIEVLGSTRYPFQNFLOOFKPL
SKTLADESIARGCVASAMVSKHVKSECEVETHPINVEWHINIASNWEKL
YVRDLSKRRKVIPIPEKGIKLTAYENTPDLPSNCIELGSCIVKINENKDIYES
HMTFTFNSDTFTFLGAQVTKSVIKSDEKKADDTEDGKKDAKQEQDNDQ
TDNNMNEKDTNDKKEKNNTSPNTELEKGEKVOCYTTPIETLLAOGSSYK
DIENFSEOIINNOHSDILEGEBRLKHNLETLIYRSRLNGICVNDDEDRILL
SLDDVENMLIYDNEKKNFIRKKEIRDLKNIYOKFVYNSKQNLGNITNHNTO
ITQCSNKPSESONIINRTKFLNNINSIQEQKNKPIEPVYTLNDIIEAENETQ
IAQKFSKLEVELEAKQAKOKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKE
18610..18772
/rpt_family="microsatellite 2E10"
/rpt_unit=ta
complement(19529..20515)
/gene="Cg3"
complement(19529..20515)
/gene="Cg3"
/codon_start=1
/evidence=experimental
/product="Cg3"
/protein_id="AAC47844.1"

```

```

/db_xref="GI:2642502"
/translation="MKNVLLNSLNKICNKKKNYNFNVLLIRNIVYDTSNRYKRE
NIFYKSEIRNDYSMSKRYVEDENDNKKKKKNIPLTWCLGINSITLIPALYLYL
LONKRRKKRGKGTMTESICPLIGDFLIRHNGIYNKSEKKEFCILYGFPCP
DICPOLKQOTTYTEKTHAKYGDITIPITISDPONDYQAQNIYKSPKPLIGLIG
TEELIKHVAKLFRVYNNVNTVNSKRENSISKNNYVNLIDHSITLHYLDITNGN
FLDFGKNATISPMVDKISEYIDDHNNKHKDPK10LAVKHKKKKKKKKK"
23488..26557
/gene="crt"
join(23488..23578,23748..24016,24180..24352,24522..24654,
24816..24887,25011..25086,25192..25274,25401..25451,
25389..25645,25788..25860,26075..26119,26257..26311,
26481..26557)
/gene="crt"
/note="PCRRT; detected by antibody"
/codon_start=1
/product="putative chloroquine resistance transporter"
/protein_id="AF26926.1"
/db_xref="GI:6724281"
/translation="MKFASKKNQKNSKNDERYRLDNLVQGNISRLGSGCLGKC
AHVFKLFEIKDNIRFIYILSTIYLSVIEITFAKRTLNKIGNYSFVSETHNFCM
IMFFIYSLFGKNGSKSRHSFNQFPAISMLDACSILAFIGLRTTGNIGSVL

```

Query Match 4.2%; Score 21; DB 3; Length 47573;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 175 gaaaaacttatatacatt 195
Db 44443 GAAAAACTTATATACATT 44423

RESULT 11
AL157832/C
LOCUS AL157832
DEFINITION Human DNA sequence from clone RP11-109A6 on chromosome 10, complete
ACCESSION AL157832
VERSION AL157832.13 GI:14625523
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 130192)
Almeida, J.
Direct Submission
Submitted (04-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 6, 2001 this sequence version replaced gi:14575107.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TrEMBL; Wp: WormPep; Information on the WormPep
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10

RP11-109A6 is from the library RPC1-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-109A6. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true right end of clone RP11-109A6 is at 130192 in this sequence. The true left end of clone RP11-218C11 is at 69543 in this sequence. The true right end of clone RP11-234G16 is at 2000 in this sequence.

FEATURES

```
source
    location/Qualifiers
    1..130192
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="10"
    /clone="RP11-109A6"
    /clone_lib="RPC1-11.1"
    1..321
    /note="L1ME3 repeat: matches 5718..6031 of consensus
    L1ME3 repeat: matches 5718..6031 of consensus"
    655..909
    /note="L1MB3 repeat: matches 5893..6145 of consensus
    L1MB3 repeat: matches 5893..6145 of consensus"
    910..1194
    /note="Aluub repeat: matches 3..288 of consensus
    Aluub repeat: matches 3..288 of consensus"
    1195..1233
    /note="L1MB3 repeat: matches 6145..6183 of consensus
    L1MB3 repeat: matches 6145..6183 of consensus"
    1234..1358
    /note="L2 repeat: matches 2618..2742 of consensus
    L2 repeat: matches 2618..2742 of consensus"
    1370..1624
    /note="L1MC/D repeat: matches 5198..5398 of consensus
    L1MC/D repeat: matches 5198..5398 of consensus"
    1680..2132
    /note="L1MC/D repeat: matches 4649..5125 of consensus"
    2133..2433
    /note="Aluud repeat: matches 1..298 of consensus
    Aluud repeat: matches 1..298 of consensus"
    2434..2452
    /note="L1MC/D repeat: matches 4632..4649 of consensus"
    2654..2838
    /note="L2 repeat: matches 2320..2502 of consensus
    L2 repeat: matches 2320..2502 of consensus"
    3096..3180
    /note="L1M4 repeat: matches 3184..3266 of consensus"
    3096..3357
    /note="L1M4 repeat: matches 3184..3479 of consensus"
    3349..3562
    /note="L1M4 repeat: matches 4741..4957 of consensus
    L1M4 repeat: matches 4741..4957 of consensus"
    3563..3905
    /note="TH1B repeat: matches 1..364 of consensus
    TH1B repeat: matches 1..364 of consensus"
    3906..4161
    /note="L1M4 repeat: matches 4957..5204 of consensus
    L1M4 repeat: matches 4957..5204 of consensus"
    4168..4445
    /note="Aluub repeat: matches 1..280 of consensus
    Aluub repeat: matches 1..280 of consensus"
    4469..4576
    /note="L1M4 repeat: matches 5191..5310 of consensus"
    7370..7726
    /note="MLT1A1 repeat: matches 1..365 of consensus
    MLT1A1 repeat: matches 1..365 of consensus"
    11608..11650
    /note="L2 repeat: matches 2583..2625 of consensus"
    12483..12566
    /note="L1MC4 repeat: matches 7896..7976 of consensus
    L1MC4 repeat: matches 7896..7976 of consensus"
    12774..12929
    repeat_region
```

```
repeat_region
    /note="L1MD3 repeat: matches 7405..7535 of consensus
    L1MC4 repeat: matches 7483..7655 of consensus"
    14386..14629
    /note="MLT1H repeat: matches 55..315 of consensus"
    14386..14508
    /note="MLT1H repeat: matches 55..185 of consensus"
    15515..15607
    /note="MLT1-INTERNAL repeat: matches 766..858 of
    consensus"
    16214..16509
    /note="MLT1H repeat: matches 51..344 of consensus
    MLT1H repeat: matches 51..344 of consensus"
    16552..16670
    /note="MLT1H repeat: matches 445..563 of consensus"
    16584..16663
    /note="MLT1H repeat: matches 456..535 of consensus"
    16680..16875
    /note="Aluux repeat: matches 30..225 of consensus"
    16680..16930
    /note="Aluux repeat: matches 30..309 of consensus"
    16939..18028
    /note="L1M4 repeat: matches 3237..4327 of consensus"
    L1M4 repeat: matches 3237..4327 of consensus"
    18030..18355
    /note="L1MD3 repeat: matches 7416..7739 of consensus
    L1MD3 repeat: matches 7416..7739 of consensus"
    18553..18708
    /note="L1ME2 repeat: matches 5167..5313 of consensus
    L1ME2 repeat: matches 5167..5313 of consensus"
    18709..18837
    /note="FLAM C repeat: matches 1..133 of consensus"
    18838..19420
    /note="L1ME2 repeat: matches 5313..5944 of consensus
    L1ME2 repeat: matches 5313..5944 of consensus"
    19419..19804
    /note="L1MB5 repeat: matches 5683..6073 of consensus"
    19419..19814
    /note="L1MB3 repeat: matches 5673..6073 of consensus"
    19815..19895
    /note="L1ME2 repeat: matches 6039..6122 of consensus
    L1ME2 repeat: matches 6039..6122 of consensus"
    22588..22685
    /note="L1MC5 repeat: matches 7800..7901 of consensus
    L1MC5 repeat: matches 7800..7901 of consensus"
    22715..23147
    /note="L1ME2 repeat: matches 5711..6143 of consensus
    L1ME2 repeat: matches 5711..6143 of consensus"
    23171..24165
    /note="L1M4 repeat: matches 3654..4703 of consensus"
    23171..24090
    /note="L1M4 repeat: matches 3760..4703 of consensus"
    25400..25587
    /note="L2 repeat: matches 2543..2750 of consensus
    L2 repeat: matches 2543..2750 of consensus"
    25673..25849
    /note="MIR repeat: matches 16..207 of consensus
    MIR repeat: matches 16..207 of consensus"
    28017..28126
    /note="11 copies 10 mer ggaagaggga 63% conserved"
    28029..28132
    /note="26 copies 4 mer ggag 62% conserved"
    28818..28957
    /note="70 copies 2 mer gt 62% conserved"
    28819..28950
    /note="33 copies 4 mer tggt 62% conserved"
    28849..28950
    /note="3 copies 34 mer 80% conserved"
    31353..32326
    /note="L1PB2 repeat: matches 5174..6152 of consensus
    L1PB2 repeat: matches 5174..6152 of consensus"
    32419..32611
    /note="Charlie1 repeat: matches 2498..2729 of consensus"
    32612..32900
    repeat_region
```

```

/Note="AlusX repeat: matches 1. .304 of consensus
AlusX repeat: matches 1. .304 of consensus"
repeat_region 32901. 32914
/Note="Charliel repeat: matches 2488. .2498 of consensus"
repeat_region 32906. 32950
/Note="Charliel repeat: matches 1822. .1868 of consensus
Charliel repeat: matches 1822. .1868 of consensus"
repeat_region 34137. 34372
/Note="LIME repeat: matches 5610. .5858 of consensus
LIME3A repeat: matches 5610. .5858 of consensus"
repeat_region 35046. 35354
/Note="AlusX repeat: matches 1. .308 of consensus
AlusX repeat: matches 1. .308 of consensus"
repeat_region 37800. 37879
/Note="8 copies 10 mer cccacacac 71% conserved"

Query Match 4.2%; Score 21; DB 9; Length 130192;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 ttgtgaaatacttatata 190
Db 76089 GTTGGAAAAATCTTATATA 76069

RESULT 12
AC099718 163721 bp DNA linear HTG 18-NOV-2001
LOCUS AC099718/c Mus musculus clone RP23-368E4, WORKING DRAFT SEQUENCE, 6 unordered
DEFINITION pieces.
ACCESSION AC099718 GI:16974217
VERSION AC099718.1
KEYWORDS HTG: HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 163721)
2 (bases 1 to 163721)
Unpublished
2 (bases 1 to 163721)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Anderson,S., Barina,N., Bastien,V., Boguslavsky,L., Boukhalter,B.,
Brown,A., Camarata,J., Campiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeRellano,K., Dewar,K., Diaz,D.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heatford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Labroque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
Maclean,C., MacDonald,P., Major,J., Margulis,N., Matthews,C.,
McCarthy,M., McEwan,P., McEneaney,K., McPherson,R., Meldrum,J.,
Menues,L., Minova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norman,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Risse,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggilo,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

DIRECT SUBMISSION
Submitted (18-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIBR

```

```

Web site: http://www.seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L17683
Center clone name: 368_E_4
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 161403 bases at least Q40
Consensus quality: 162593 bases at least Q30
Consensus quality: 162938 bases at least Q20
Insert size: 163000; agarose-fp
Insert size: 163221; sum-of-ctrls
Quality coverage: 8.7 in Q20 bases; agarose-fp
Quality coverage: 8.7 in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1925: contig of 1925 bp in length
* 1926 2025: gap of 100 bp
* 2026 2677: contig of 652 bp in length
* 2678 2777: gap of 100 bp
* 2778 3978: contig of 1201 bp in length
* 3979 4078: gap of 100 bp
* 4079 25935: contig of 21857 bp in length
* 25936 26035: gap of 100 bp
* 26036 132236: contig of 106201 bp in length
* 132237 132336: gap of 100 bp
* 132337 163721: contig of 31385 bp in length.
Location/Qualifiers
1. 163721
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-368E4"
/clone_lib="RP23-368E4 Female Mouse BAC"
1. 1925
/Note="assembly-fragment"
clone_end:sp6
vector_side:left"
misc_feature 2026. 2677
/Note="assembly-fragment"
2778. 3978
/Note="assembly-fragment"
4079. 25935
/Note="assembly-fragment"
26036. 132236
/Note="assembly-fragment"
132337. 163721
/Note="assembly-fragment"
clone_end:r7
vector_side:right"
BASE COUNT 51477 a 30666 c 30186 g 50867 t 525 others
ORIGIN
Query Match 4.2%; Score 21; DB 2; Length 163721;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 375 ttgtcctaataatgatgc 395
Db 150753 TTGTCTTAAAAATTTGATG 150733

RESULT 13
AC018633/c

```

LOCUS AC018633 169234 bp DNA linear PRI 21-JAN-2000
 DEFINITION Homo sapiens clone UMGc:djs1 or RP11-16G1 from 7p14-15, complete
 sequence.
 ACCESSION AC018633
 VERSION AC018633.2 GI:6729063
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 169234)
 Kaul, R.K., Yu, D., Wong, G.K.-S., Magness, C.L., Green, E.D., Green, P.
 and Olson, M.V.
 Large-scale MCD Mapping and Sequencing of Human Chromosome 7
 Unpublished
 2 (bases 1 to 169234)
 Bubb, K.L., Desmarais, C.L., Ramsey, S.A. and Hubley, R.M.
 Direct Submision
 Submitted (15-DEC-1999) Human Genome Center, University of
 Washington, Box 352145, Seattle, WA 98195, USA
 3 (bases 1 to 169234)
 Kaul, R.K. and Richards, B.K.
 Direct Submision
 Submitted (21-JAN-2000) Human Genome Center, University of
 Washington, Box 352145, Seattle, WA 98195, USA
 On Jan 21, 2000 this sequence version replaced gi:6579285.

COMMENT

----- Genome Center:

Center: University of Washington Genome Center

Center code: UMGc

Web site: http://genome.washington.edu

Contact: umgchelp@u.washington.edu

----- Project Information

Center project name: Chr-7

Center clone name: djs1 (RP11-16G1)

----- Summary Statistics

Sequencing vector: M13; 100% of reads

Chemistry: Dye-primer Bodipy; 90% of reads

Chemistry: Dye-terminator Big Dye; 10% of reads

Assembly program: Phrap; version 0.990319

Insert size: 169 234; sum-of-ctrligs

Quality coverage: 11.1X in Q20 bases; sum-of-ctrligs

----- Overlapping Sequences:

5': UMGc:djs21

3': UMGc:djs29

----- Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than

1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the

Genbank flat file format but are available as part

of this entry's ASN.1 file.

Double stranded (DS) coverage: 96.4%

DS or two chemistry coverage: 97.1%

Single stranded regions: 10

----- Sequence Validation:

This sequence has been validated by Multiple Complete Digest

fingerprinting. Comparison of the experimentally derived digest

fragments with sequence-predicted fragments is given below.

The electronically-digested sequence consists of both insert and

vector, in order to accurately represent the entire circular BAC.

Small fragments below a variable cutoff (approximately 400-600 bp)

are not resolved in the fingerprint and hence do not appear

in the table. There are no significant remaining discrepancies

between the experimental and predicted values. Uniquely ordered

fragments are separated by dashed lines.

BgIII

EcoRI

NsiI

FP	Seq	FP	Seq	FP	Seq
2055.00	2061.00	8954.00	8687.00	11080.00	11109.00
5487.00	5270.00	7648.00	7660.00	4611.00	4562.00
1913.00	1893.00	7648.00	7556.00	11080.00	11183.00
1508.00	1491.00	838.00	801.00	1377.00	1369.00
12901.00	13322.00	4399.00	4303.00	2193.00	2193.00
1004.00	992.00	4140.00	4094.00	3735.00	3692.00
529.00	503.00	660.00	643.00	614.00	599.00
10358.00	10526.00	3053.00	3100.00	3918.00	3851.00
3805.00	3819.00	5075.00	5106.00	614.00	593.00
1674.00	1645.00	6909.00	6771.00	2570.00	2518.00
2228.00	2170.00	1911.00	1874.00	1997.00	1995.00
3051.00	2974.00	3635.00	3584.00	6199.00	6176.00
14433.00	14779.00	3975.00	3996.00	4913.00	5009.00
1508.00	1475.00	897.00	864.00	2193.00	2184.00
1865.00	1874.00	1410.00	1399.00	12585.00	12586.00
2228.00	2269.00	5716.00	5790.00	1819.00	1821.00
2055.00	2043.00	3306.00	3304.00	709.00	690.00
1347.00	1325.00	3975.00	3893.00	6528.00	6484.00
4492.00	4449.00	897.00	876.00	1944.00	1934.00
3805.00	3757.00	7648.00	7475.00	3127.00	3138.00
3051.00	3031.00	3635.00	3611.00	16492.00	16238.00
1674.00	1676.00	838.00	824.00	2842.00	2741.00
10030.00	10132.00	3053.00	2987.00	6528.00	6572.00
4492.00	4478.00	10778.00	10756.00	10111.00	9669.00
3278.00	3209.00	1250.00	1234.00	11366.00	11758.00
1347.00	1339.00	2135.00	2108.00	4059.00	3979.00
8691.00	8644.00	5075.00	4993.00	1997.00	1990.00
1347.00	1331.00	897.00	875.00	2570.00	2445.00
4350.00	4341.00	1103.00	1082.00	2570.00	2576.00
12901.00	12601.00	3635.00	3602.00	11366.00	11207.00
3278.00	3339.00	11685.00	11548.00	10111.00	10101.00
12901.00	13224.00	1410.00	1385.00	13982.00	13668.00
3051.00	3072.00	2819.00	2826.00	-----	-----
5487.00	5453.00	19500.00	19348.00	-----	-----
2076.00	2128.00	4140.00	4057.00	-----	-----
1535.00	1554.00	2745.00	2716.00	-----	-----

```

-----
3051.00 3063.00 1734.00 1710.00
4775.00 4741.00 3053.00 3104.00
10030.00 10021.00 8954.00 8789.00
-----
1147.00 1151.00
2285.00 2234.00
897.00 894.00
688.00 677.00
966.00 955.00
688.00 669.00
-----
FEATURES
source      Location/Qualifiers
1.169234
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7p14-15"
/clone="djs1 (RP11-1661)"
/cell_line="Male Blood"
/clone_lib="RPC-11 Human Male BAC Library"
392..494
/rpt_family="MER21"
complement(1075..1364)
/rpt_family="Alu"
complement(2210..2373)
/standard_name="SWS2626"
/feature="GenBank Accession Number: G12923"
repeat_region
2594..2876
/rpt_family="Alu"
complement(3252..3702)
/rpt_family="MLT1"
complement(4489..4790)
/rpt_family="Alu"
complement(5307..5594)
/rpt_family="Alu"
5847..6152
/rpt_family="Alu"
complement(6970..7082)
/standard_name="SWS1000"
/feature="GenBank Accession Number: G00158"
repeat_region
complement(7545..7634)
/rpt_family="Alu"
complement(8959..9186)
/rpt_family="Alu"
10137..10423
/rpt_family="Alu"

```

```

Query Match      4.2%; Score 21; DB 9; Length 169234;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 177 aaatcttatatatttt 197
|||||
Db 133656 AAAATCTTATATACATTTT 133636

RESULT 14
LOCUS AF034998 1295 bp mRNA linear INV 26-JAN-1999
DEFINITION Hyphantria cunea immune-related Hdd1 mRNA, complete cds.
ACCESSION AF034998
VERSION AF034998.1 GI:4090963
KEYWORDS fall webworm moth.
SOURCE ORGANISM Hyphantria cunea

```

```

REFERENCE
AUTHORS Shin,S.W., Park,S.-S., Park,D.S., Kim,M.G., Kim,S.C., Brey,P.T. and Park,H.Y.
TITLE Isolation and characterization of immune-related genes from the fall webworm, Hyphantria cunea, using PCR-based differential display and subtractive cloning
JOURNAL Insect Biochem. Mol. Biol. 28 (11), 827-837 (1998)
MEDLINE 99035790
REFERENCE 2 (bases 1 to 1295)
AUTHORS Shin,S.W., Park,S.-S. and Park,H.-Y.
TITLE Direct Submission
JOURNAL Submitted (19-NOV-1997) Insect Resources Lab., Korea Research Institute of Bioscience and Biotechnology, P.O. Box 115, Yuseong, Taejeon 305-600, Korea

FEATURES
source      Location/Qualifiers
1.1295
/organism="Hyphantria cunea"
/db_xref="taxon:39466"
6..968
/codon_start=1
/product="Immune-related Hdd1"
/protein_id="AAD09279.1"
/db_xref="GI:4090964"
/translation="MYVLAIVGLVSTASAKDITVGRGGSLKRFQNTASPTIMROA
KTVFNVSDREVITVOYAIVDRREDKGEAVITIEGNGHKVITELKSPAVRFDEI
RVYAKOEDDIQSPVDGQOONIQPFQDIOVNVKQVQVAKVDTGSPVSTQIPN
GDRKVOHPVVIQGEDNQVARNPDIDIGQETQITQVPEELQSKNQNOSDIABE
QNKEMQAVIDVGFQNKPGESKKPTAEPMKPIQVADHPFGSKLDINIDHVRVIGIDG
QKVQSTENKVKQGOQEKSIQVOQTNDISOAKQVNVPIPLKQ"

BASE COUNT 472 a 246 c 255 g 322 t
ORIGIN

```

```

Query Match      4.0%; Score 20; DB 3; Length 1295;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 194 ttttcagagattgacctt 213
|||||
Db 286 TTTTCAGAGATTTGACTTT 305

RESULT 15
LOCUS AC107995 57205 bp DNA linear HTG 24-JAN-2002
DEFINITION Homo sapiens chromosome 15 clone RP11-577014 map 15, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC107995
VERSION AC107995.1 GI:18308723
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 57205)
Birten,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-577014
Unpublished
2 (bases 1 to 57205)
Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campolano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Huime,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Lahocque,K., Lamazares,R.,
Landers,T., Lehoczeky,J., Levine,R., Liu,G., Maclean,C.,
Macdonald,P., Major,J., Margulis,N., Matthews,C., McCarthy,M.,

```

TITLE
JOURNAL
COMMENT

McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T.,
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J.,
Peterson, K., Phunhthang, P., Pierre, N., Pollara, V., Raymond, C.,
Rettl, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliou, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WtBR

Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L24565
Center clone name: 577_O_14

NOTE: This record contains 69 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 706: contig of 706 bp in length
* 707 806: gap of 100 bp
* 807 1528: contig of 722 bp in length
* 1529 1628: gap of 100 bp
* 1629 2319: contig of 691 bp in length
* 2320 2419: gap of 100 bp
* 2420 3128: contig of 709 bp in length
* 3129 3228: gap of 100 bp
* 3229 3949: contig of 721 bp in length
* 3950 4049: gap of 100 bp
* 4050 4784: contig of 735 bp in length
* 4785 4884: gap of 100 bp
* 4885 5614: contig of 730 bp in length
* 5615 5714: gap of 100 bp
* 5715 6425: contig of 711 bp in length
* 6426 6525: gap of 100 bp
* 6526 7261: contig of 736 bp in length
* 7262 7361: gap of 100 bp
* 7362 8097: contig of 736 bp in length
* 8098 8197: gap of 100 bp
* 8198 8918: contig of 721 bp in length
* 8919 9018: gap of 100 bp
* 9019 9757: contig of 739 bp in length
* 9758 9857: gap of 100 bp
* 9858 10593: contig of 736 bp in length
* 10594 10693: gap of 100 bp
* 10694 11441: contig of 748 bp in length
* 11442 11541: gap of 100 bp
* 11542 12280: contig of 739 bp in length
* 12281 12380: gap of 100 bp
* 12381 13110: contig of 730 bp in length
* 13111 13210: gap of 100 bp
* 13211 13922: contig of 712 bp in length
* 13923 14022: gap of 100 bp
* 14023 14727: contig of 705 bp in length
* 14728 14827: gap of 100 bp

14828 15553: contig of 726 bp in length
* 15554 15653: gap of 100 bp
* 15654 16377: contig of 723 bp in length
* 16377 16476: gap of 100 bp
* 16477 17185: contig of 709 bp in length
* 17186 17285: gap of 100 bp
* 17286 18035: contig of 750 bp in length
* 18036 18135: gap of 100 bp
* 18136 18872: contig of 737 bp in length
* 18873 18972: gap of 100 bp
* 18973 19721: contig of 749 bp in length
* 19722 19821: gap of 100 bp
* 19822 20542: contig of 721 bp in length
* 20543 20642: gap of 100 bp
* 20643 21361: contig of 719 bp in length
* 21362 21461: gap of 100 bp
* 21462 22213: contig of 752 bp in length
* 22214 22313: gap of 100 bp
* 22314 23036: contig of 723 bp in length
* 23037 23136: gap of 100 bp
* 23137 232861: contig of 725 bp in length
* 232862 23961: gap of 100 bp
* 23962 24713: contig of 752 bp in length
* 24714 24813: gap of 100 bp
* 24814 25557: contig of 744 bp in length
* 25558 25657: gap of 100 bp
* 25658 26405: contig of 748 bp in length
* 26406 26505: gap of 100 bp
* 26506 27257: contig of 752 bp in length
* 27258 27357: gap of 100 bp
* 27358 28094: contig of 737 bp in length
* 28095 28194: gap of 100 bp
* 28195 28929: contig of 735 bp in length
* 28930 29029: gap of 100 bp
* 29030 29749: contig of 720 bp in length
* 29750 29849: gap of 100 bp
* 29850 30595: contig of 746 bp in length
* 30596 30695: gap of 100 bp
* 30696 31408: contig of 713 bp in length
* 31409 31508: gap of 100 bp
* 31509 32221: contig of 713 bp in length
* 32222 32321: gap of 100 bp
* 32322 33304: contig of 723 bp in length
* 33045 33144: gap of 100 bp
* 33145 33893: contig of 749 bp in length
* 33894 33993: gap of 100 bp
* 33994 34717: contig of 724 bp in length
* 34718 34817: gap of 100 bp
* 34818 35549: contig of 732 bp in length
* 35550 35649: gap of 100 bp
* 35650 36353: contig of 704 bp in length
* 36354 36453: gap of 100 bp
* 36454 37182: contig of 729 bp in length
* 37183 37282: gap of 100 bp
* 37283 37998: contig of 716 bp in length
* 37999 38098: gap of 100 bp
* 38099 38849: contig of 751 bp in length
* 38850 38949: gap of 100 bp
* 38950 39699: contig of 750 bp in length
* 39700 39799: gap of 100 bp
* 39800 40533: contig of 734 bp in length
* 40534 40633: gap of 100 bp
* 40634 41360: contig of 727 bp in length
* 41361 41460: gap of 100 bp
* 41461 42209: contig of 749 bp in length
* 42210 42309: gap of 100 bp
* 42310 43061: contig of 752 bp in length
* 43062 43161: gap of 100 bp
* 43162 43892: contig of 731 bp in length
* 43893 43992: gap of 100 bp
* 43993 44724: contig of 732 bp in length
* 44725 44824: gap of 100 bp
* 44825 45545: contig of 721 bp in length

```
* 45546 45645: gap of 100 bp
* 45646 46393: contig of 748 bp in length
* 46394 46493: gap of 100 bp
* 46494 47186: contig of 693 bp in length
* 47187 47286: gap of 100 bp
* 47287 48011: contig of 725 bp in length
* 48012 48111: gap of 100 bp
* 48112 48835: contig of 724 bp in length
* 48836 48935: gap of 100 bp
* 48936 49667: contig of 732 bp in length
* 49668 49767: gap of 100 bp
* 49768 50519: contig of 752 bp in length
* 50520 50619: gap of 100 bp
* 50620 51350: contig of 731 bp in length
* 51351 51450: gap of 100 bp
* 51451 52179: contig of 729 bp in length
* 52180 52279: gap of 100 bp
* 52280 52997: contig of 718 bp in length
* 52998 53097: gap of 100 bp
* 53098 53825: contig of 728 bp in length
* 53826 53925: gap of 100 bp
* 53926 54648: contig of 723 bp in length
* 54649 54748: gap of 100 bp
* 54749 55501: contig of 753 bp in length
* 55502 55601: gap of 100 bp
* 55602 56350: contig of 749 bp in length
* 56351 56450: gap of 100 bp
* 56451 57205: contig of 755 bp in length.
```

Query Match 4.0%; Score 20; DB 2; Length 57205;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 gctgcagctctaaccag 47
|||||
Db 6794 GCTGCAGCTTCAATCCAG 6813

Search completed: May 22, 2002, 08:30:10
Job time: 10700 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 06:46:20 ; Search time 373 Seconds
(without alignments)
2301.494 Million cell updates/sec

Title: US-09-693-205-7_COPY_1_500

Perfect score: 500
Sequence: 1 atgatttcacgaagacatc.....gcagaattctgtatcaaa 500

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

N_Geneseq_032802.*
1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	500	100.0	12792	22	AAH20176
2	500	100.0	12793	22	AAH20174
3	500	100.0	12793	22	AAH20178
4	500	100.0	12793	22	AAH20179
5	500	100.0	12793	22	AAH20182
6	35	7.0	11453	22	AAH20175
7	23	4.6	23	22	AAH20130
8	19	3.8	721	16	AAH35105
9	19	3.8	1543	23	AAH83269

10	19	3.8	1588	22	AAH72857
C 11	19	3.8	3365	18	AAH74557
C 12	19	3.8	4057	23	ABL14512
C 13	19	3.8	7076	24	ABL14145
C 14	18	3.6	431	22	AAH12392
C 15	18	3.6	513	22	AAH55109
C 16	18	3.6	516	22	AAH55108
C 17	18	3.6	1034	23	AAH87201
C 18	18	3.6	1440	22	AAH53781
C 19	18	3.6	1481	21	AAH47212
C 20	18	3.6	1489	21	AAH39202
C 21	18	3.6	3180	22	AAH54577
C 22	18	3.6	3736	20	AAH84918
C 23	18	3.6	6012	24	ABL34554
C 24	18	3.6	6203	22	AAH54575
C 25	18	3.6	6731	24	ABL33061
C 26	18	3.6	9391	22	ABH16800
C 27	18	3.6	10078	22	AAH40045
C 28	18	3.6	10078	22	AAH91462
C 29	18	3.6	11459	22	AAH06876
C 30	18	3.6	15667	24	ABL34146
C 31	18	3.6	67087	21	AAH22280
C 32	18	3.6	100848	22	AAH28552
C 33	17	3.4	51	22	AAH32440
C 34	17	3.4	287	22	AAH28202
C 35	17	3.4	287	22	AAH28204
C 36	17	3.4	305	19	AAH61286
C 37	17	3.4	305	19	AAH58543
C 38	17	3.4	305	17	AAH06306
C 39	17	3.4	305	22	AAH563514
C 40	17	3.4	305	22	AAH10065
C 41	17	3.4	305	22	AAH93422
C 42	17	3.4	305	22	AAH84736
C 43	17	3.4	305	22	AAH02487
C 44	17	3.4	334	22	AAH06614
C 45	17	3.4	382	22	AAH92739

ALIGNMENTS

RESULT 1	
AAH20176	
ID	AAH20176 standard; DNA; 12792 BP.
XX	
AC	AAH20176;
XX	
DT	09-AUG-2001 (first entry)
XX	
DE	Human mutated spastin nucleotide sequence SEQ ID NO:7.
XX	
KW	Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
KW	autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KW	neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KW	reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
KW	atrophy of upper cerebellar vermis; absence of Purkinje cell;
KW	abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
XX	
OS	Homo sapiens.
XX	
XX	Synthetic.
XX	
FT	Key
FT	CDS
FT	Location/Qualifiers
XX	77..6604
XX	/*tag="a
XX	/product="mutated spastin"
XX	
XX	W0200129266-A2.
XX	
PD	26-APR-2001.
XX	
PF	20-OCT-2000; 2000WO-US29130.
XX	
PR	20-OCT-1999; 99US-0160588.

XX (UYMC-) UNIV MCGILL.
 PA (HOPI-) HOPITAL SAINTE-JUSTINE.
 XX
 XX
 PI Hudson TJ, Engert J, Richter A;
 XX
 XX WPI: 2001-308494/32.
 DR P-PSDB; AAB97821.
 DR
 XX
 XX
 PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -
 XX
 PS Claim 1; Page -; 76pp; English.

CC The present invention describes human and mouse spastin and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSAACS)) gene sequences (1). The spastin gene has been mapped to
CC chromosome 13q11. (1) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (1), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (1) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (1) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (1). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with Spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes a mutated human spastin.
CC N.B. The present sequence is not given in the present specification but
CC is derived from the human spastin nucleotide sequence (AAH20174) as
CC stated on page 14.

SQ Sequence 12792 BP; 4163 A; 2256 C; 2487 G; 3886 T; 0 other;

Query Match	100.0%	Score 500; DB 22;	Length 12792;
Best Local Similarity	100.0%;	Pred. No. 4.7e-239;	
Matches 500; Conservative	0;	Mismatches	0; Gaps 0;

Qy	1	algtatttcagaagacacatgttactaagctcgacagctcttcaatccagaaagatttcagc	60
Db	1	atgatttcagaagacacatgttactaagctcgacagctcttcaatccagaaagatttcagc	60
Qy	61	tctttcaagaagaatgaatgaatactctcgccgcgcagagaagaatttgcttcaatgtrt	120
Db	61	tctttcaagaagaatgaatgaatactctcgccgcgcagagaagaatttgcttcaatgtrt	120
Qy	121	tccattttgaataaacagaaatccaccacatctgttttcattggtcctaagaatggttttgaaaaa	180
Db	121	tccattttgaataaacagaaatccaccacatctgttttcattggtcctaagaatggttttgaaaaa	180
Qy	181	tctttataatacatcttttcagaagatttgactttattttgatbagaatgacacttatccccag	240
Db	181	tctttataatacatcttttcagaagatttgactttattttgatbagaatgacacttatccccag	240
Qy	241	aactatactagaagaagtcagacatgtgtgtgaaatccatataagaccagatcccatcgtt	300
Db	241	aactatactagaagaagtcagacatgtgtgtgaaatccatataagaccagatcccatcgtt	300
Qy	301	agtcattttagaagatgaatctctgaagacagcttcccaagaattttttagcagaactttgaca	360
Db	301	agtcattttagaagatgaatctctgaagacagcttcccaagaattttttagcagaactttgaca	360

QY	361	aaacttggaggtgttgccttaaaaaatagatgacatcatacaatccgttttaa	420
Db	361	aaacttggaggtgttgccttaaaaaatagatgacatcatacaatccgttttaa	420
QY	421	aaaataatttcacacattaccaaatgctgtttgcataatatggagaagtgcatt	480
Db	421	aaaataatttcacacacattaccaaatgctgtttgcataatatggagaagtgcatt	480
QY	481	gcagaatctgtgtaatcaaa	500
Db	481	gcagaatctgtgtaatcaaa	500

RESULT	ID	standard; DNA; 12793 BP.
AAH20174	AAH20174	
XX	AAH20174;	
AC		
XX		
DT	09-AUG-2001	(first entry)

	Human spastin nucleotide sequence seq ID NO:1.
DE	
XX	Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
KW	autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KW	neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KW	reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
KW	atrophy of upper cerebellar vermis; absence of Purkinje cell;
KW	abnormal neuronal lipid storage; genetic disorder; characterisation; ds
XX	
OS	Homo sapiens.

EH	key	Location/Qualifiers
FT	CDS	77..11566
FT		/tag= a
FT		/product=
XX		"spastin"
XX		
PN		

WG200129265-A2.

PD 26-APR-2001.

PF 20-OCT-2000; 2000WO-US29130.

PR 20-OCT-1999; 99US-0160588.

PA (UYMC-) UNIV MCGILL.

XX

XX

DR P-PSDB; AAB97819.

PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -

PS Claim 1; Fig 9; 76pp; English.

CC The present invention describes human and mouse *spastin*, and mutated
CC human *spastin* (autosomal recessive spastic ataxia of Charlevoix-Seguenay
CC (ARSACS)) gene sequences (1). The *spastin* gene has been mapped to
CC chromosome 13q11. (1) have neuroprotective activities and can be used in
CC gene therapy and as a *spastin* polypeptide agonists. (1), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (1) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypomyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (1) can also be used to produce

CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (I). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with Spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes human spastin as given in the present invention.

XX
XX Sequence 12793 BP; 4163 A; 2256 C; 2487 G; 3887 T; 0 other;

Query Match	100.0%;	Score 500;	DB 22;	Length 12793;
Best Local Similarity	100.0%;	Pred. No. 4.7e-239;		
Matches 500;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	atgattcaaggaagaccatctgactctgacgcgcgaagctcttaaacccgaagatttcgacg	60
Db	1	atgattcttaacgaaggaaccacagtaactcagctcgacgctcttaaacccgaagatttcgacg	60
QY	61	tcttatcaaggaagtaatgtgatacatctctgacctgcgcagagaattgtattgtcaatgtta	120
Db	61	tcttatcaaggaagtaatgtgatacatctctgacctgcgcagagaattgtattgtcaatgtta	120
QY	121	tccatttgatgaaacacgaataatcccatctgtttctatgtgcttaagaattgttggaaaaa	180
Db	121	tccatttgatgaaacacgaataatcccatctgtttctatgtgcttaagaattgttggaaaaa	180
QY	181	tcttatatacatctcttcacagagatttcgacttctatgttgatgagatgcacttaaccag	240
Db	181	tcttatatacatctcttcacagagatttcgacttctatgttgatgagatgcacttaaccag	240
QY	241	aactatctctgaggaagaagtcgaagaattgtggaactcttaataagactcaggaattccatgct	300
Db	241	aactatctctgaggaagaagtcgaagaattgtggaactcttaataagactcaggaattccatgct	300
QY	301	agtcattttagacgaatgaatcttgagaagcagcttcacgaatttcttagcaacacttgaca	360
Db	301	agtcattttagacgaatgaatcttgagaagcagcttcacgaatttcttagcagaacttgaca	360
QY	361	aaaactctggagggcttctccttaaaaaattagatgacatcatatacaaatccgcttataa	420
Db	361	aaaactctggagggcttctccttaaaaaattagatgacatcatatacaaatccgcttataa	420
QY	421	aaaataatcatccacatcatccaagtgctgttttcagataatggaagagatgcccatt	480
Db	421	aaaataatcatccacatcatccaagtgctgttttcagataatggaagagatgcccatt	480
QY	481	gcagaaattgtgtaatcaaa 500	
Db	481	gcagaaattgtgtaatcaaa 500	

RESULT	3
AAH20178	
ID	AAH20178 standard; DNA; 12793 BP.

AAH20178; AC

DT 09-AUG-2001 (first entry)

Human mutated spastin nucleotide sequence SEQ ID NO:11.

KW Human, mouse; spastin; ARSACS; chromosome 13q11; identification;
KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KW reduced motor nerve velocity; hypomyelination of retinal nerve fibre;
KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds

✕

OS	Homo sapiens.
OS	Synthetic.
XX	
PN	WO200129266-A2.

PD 26-APR-2001.

PF 20-OCT-2000; 2000WO-US29130.

PR 20-OCT-1999; 99US-0160588.

PA (UYMC-) UNIV MCGILL.

XX

XX

[illegible]

PT useful for diagnosing autosomal recessive spastic ataxia of

PT gene sequence -

PS Claim 1; Page -; 76pp; English.

CC The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay

CC (MNA3C3) gene sequences (1). The spactin gene has been mapped to
CC chromosome 13q11. (I) have neuroprotective activities and can be used

or their complements can be useful for assaying the presence of a nucle

diagnosis of an early onset neurodegenerative disease in an individual.

reduced motor nerve velocity, hypermyelination of retinal nerve fibres,

abnormal neuronal lipid storage. (I) can also be used to produce

CC markers, to identify genetic disorders, as hybridisation probes or

CC analysis, characterisation or therapeutic use, or as markers for tissue

CC the present invention can be used to identify subjects having or at risk

CC or activity of (I). The assays can be utilised to identify a subject

CC protein or spastin gene expression or activity. The present sequence

CC N.B. The present sequence is not given in the present specification but

CC stated on page 14.

Sequence 12793 BP; 4163 A; 2255 C; 2487 G; 3888 T; 0 other;

Best Local Similarity 100.08; Pred. No. 4.7e-239;

[illegible]

A vertical ruler scale with markings from 0 to 10 cm. The markings are in centimeters, with millimeter increments indicated between the centimeter numbers. The scale is oriented vertically, with 0 at the top and 10 at the bottom.

[illegible]

Db 181 tcttatatacatlcttcagaggattgacttattatgatgagatgccactatccccag 240
 QY 241 aactatactagagaaagtcagacatgtgtggaactcattagactcagagttccatcgtt 300
 Db 241 aactatactagagaaagtcagacatgtgtggaactcattagactcagagttccatcgtt 300
 QY 301 agtcattttagacgatgaatcctgaagcacagcttccagaaatttttagcagacattgtaca 360
 Db 301 agtcattttagacgatgaatcctgaagcacagcttccagaaatttttagcagacattgtaca 360
 QY 361 aaactctggagggttgccttaaaaaattagatgcatctatacaaatccgcttattaa 420
 Db 361 aaactctggagggttgccttaaaaaattagatgcatctatacaaatccgcttattaa 420
 QY 421 aaatatattcattccaccattaccacaaagtgtgttttcagagataatggagaagatgcatt 480
 Db 421 aaatatattcattccaccattaccacaaagtgtgttttcagagataatggagaagatgcatt 480
 QY 481 gcagaaatgtgtatcaaa 500
 Db 481 gcagaaatgtgtatcaaa 500

RESULT 4
 ID AAH20179 standard; DNA; 12793 BP.
 AC AAH20179;
 XX
 DT 09-AUG-2001 (first entry)
 XX
 DE Human mutated spastin nucleotide sequence SEQ ID NO:12.
 XX
 KW Human: mouse: spastin; ARSACS; chromosome 13q11; identification:
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation:
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis:
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO200129266-A2.
 XX
 PD 26-APR-2001.
 XX
 PF 20-OCT-2000; 2000MO-US29130.
 XX
 PR 20-OCT-1999; 99GS-0160588.
 XX
 PA (UYMC-) UNIV MCGILL.
 PA (HOP-) HOPITAL SAINTE-JUSTINE.
 XX
 PI Hudson TJ, Engert J, Richter A;
 XX
 DR WPI; 2001-308494/32.
 XX
 PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -
 XX
 PS Claim 1; Page -: 76pp; English.
 XX
 CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.

CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with spastin
 CC protein or spastin gene expression or activity. The present sequence
 CC encodes a mutated human spastin.
 CC N.B. The present sequence is not given in the present specification but
 CC is derived from the human spastin nucleotide sequence (AAH20174) as
 CC stated on page 14.
 XX
 SQ Sequence 12793 BP; 4163 A; 2257 C; 2487 G; 3886 T; 0 other;

Query Match 100.0%; Score 500; DB 22; Length 12793;
 Best Local Similarity 100.0%; Pred. No. 4,7e-239;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgattacagagaagaccatgtactacagctgcagcttcaatccagaagattgcagc 60
 Db 1 atgattacagagaagaccatgtactacagctgcagcttcaatccagaagattgcagc 60
 QY 61 tcttatcaaggaagtaatacatcttgccttgcagagaattgatttcaatgta 120
 Db 61 tcttatcaaggaagtaatacatcttgccttgcagagaattgatttcaatgta 120
 QY 121 tccatttgatgaagaacagaatacccatctgtttatagtgcttaagatggttggaaaaa 180
 Db 121 tccatttgatgaagaacagaatacccatctgtttatagtgcttaagatggttggaaaaa 180
 QY 181 tcttatatacatlcttcagaggattgacttattatgatgagatgccacttaccag 240
 Db 181 tcttatatacatlcttcagaggattgacttattatgatgagatgccacttaccag 240
 QY 241 aactatactagagaaagtcagacatgtgtggaactcattagactcagagttccatcgtt 300
 Db 241 aactatactagagaaagtcagacatgtgtggaactcattagactcagagttccatcgtt 300
 QY 301 agtcattttagacgatgaatcctgaagcacagcttccagaaatttttagcagacattgtaca 360
 Db 301 agtcattttagacgatgaatcctgaagcacagcttccagaaatttttagcagacattgtaca 360
 QY 361 aaactctggagggttgccttaaaaaattagatgcatctatacaaatccgcttattaa 420
 Db 361 aaactctggagggttgccttaaaaaattagatgcatctatacaaatccgcttattaa 420
 QY 421 aaatatattcattccaccattaccacaaagtgtgttttcagagataatggagaagatgcatt 480
 Db 421 aaatatattcattccaccattaccacaaagtgtgttttcagagataatggagaagatgcatt 480
 QY 481 gcagaaatgtgtatcaaa 500
 Db 481 gcagaaatgtgtatcaaa 500

RESULT 5
 ID AAH20182 standard; DNA; 12793 BP.
 AC AAH20182;
 XX
 DT 09-AUG-2001 (first entry)
 XX
 DE Human mutated spastin nucleotide sequence SEQ ID NO:15.
 XX

KW Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
KM autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KM neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KM reduced motor nerve velocity; hypermyelination of retinal nerve fibres;
KM atrophy of upper cerebellar vermis; absence of Purkinje cell;
KM abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 77..11566
FT /*tag- a
FT /product= "mutated spastin"
XX
PD WO200129266-A2.
XX 26-APR-2001.
XX
PF 20-OCT-2000; 2000MO-US29130.
XX
PR 20-OCT-1999; 99US-0160588.
XX
PA (UYMC-) UNIV MCGILL.
XX (HOP1-) HOPITAL SAINTE-JUSTINE.
PI Hudson TJ, Engert J, Richter A;
XX WPI; 2001-308494/32.
DR P-SDB; AAB97823.
XX
PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -
XX
PS Claim 1; Page -: 76pp; English.
XX
CC The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
CC chromosome 13q11. (I) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (II), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (I) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (I). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with Spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes a mutated human spastin.
CC N.B. The present sequence is not given in the present specification but
CC is derived from the human spastin nucleotide sequence (AAH20174) as
CC stated on page 14.
XX
SQ Sequence 12793 BP; 4163 A; 2257 C; 2487 G; 3886 T; 0 other;

Query Match 100.0%; Score 500; DB 22; Length 12793;
Best Local Similarity 100.0%; Pred. No. 4.7e-239;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 atgattacaggaagaccatgtaactcagctgcagcttctaataccagaagatttcacg 60

|||||
Db 1 atgattacaggaagaccatgtaactcagctgcagcttctaataccagaagatttcacg 60
Oy 61 tctatacagaagtaataacatattctgctgcgcagagaattgtttcaatgtga 120
|||||
Db 61 tctatacagaagtaataacatattctgctgcgcagagaattgtttcaatgtga 120
Oy 121 tccatttgatgaagaacagaatcacccatctgtttcatatgcttaagatgtttgaaaaa 180
|||||
Db 121 tccatttgatgaagaacagaatcacccatctgtttcatatgcttaagatgtttgaaaaa 180
Oy 181 tcttataatactttttcagaagattgactttttttatgataagttgcacattccccag 240
|||||
Db 181 tcttataatactttttcagaagattgactttttttatgataagttgcacattccccag 240
Oy 241 aactatacagaagagtcacagatctgttgaactcaatagactcaggaattccatcgtt 300
|||||
Db 241 aactatacagaagagtcacagatctgttgaactcaatagactcaggaattccatcgtt 300
Oy 301 agtcattttagaagatgaatctgagcacagcgttccagaatttttagcagacattgtaca 360
|||||
Db 301 agtcattttagaagatgaatctgagcacagcgttccagaatttttagcagacattgtaca 360
Oy 361 aaacttggaaggtttgttccttaaaaattagatgacatctatacaacatccgctattaa 420
|||||
Db 361 aaacttggaaggtttgttccttaaaaattagatgacatctatacaacatccgctattaa 420
Oy 421 aaatatatcatcattcacatcaccaagtgctgttttcagagataatggaagaatgacatt 480
|||||
Db 421 aaatatatcatcattcacatcaccaagtgctgttttcagagataatggaagaatgacatt 480
Oy 481 gcagaagaattgtgtatcaaa 500
|||||
Db 481 gcagaagaattgtgtatcaaa 500

RESULT 6
AAH20175
ID AAH20175 standard; DNM; 11493 BP.
XX
AC AAH20175;
XX
DT 09-AUG-2001 (first entry)
XX
DE Mouse spastin nucleotide sequence SEQ ID NO:3.
XX
KW Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
KM autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KM neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KM reduced motor nerve velocity; hypermyelination of retinal nerve fibres;
KM atrophy of upper cerebellar vermis; absence of Purkinje cell;
KM abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 1..11493
FT /*tag- a
FT /product= "spastin"
XX
PD WO200129266-A2.
XX
PD 26-APR-2001.
XX
PF 20-OCT-2000; 2000MO-US29130.
XX
PR 20-OCT-1999; 99US-0160588.
XX
PA (UYMC-) UNIV MCGILL.
XX (HOP1-) HOPITAL SAINTE-JUSTINE.
PI Hudson TJ, Engert J, Richter A;
XX

DR WPI; 2001-308494/32.
 DR P-PSDB; AAB97820.
 XX
 PT New isolated polynucleotide, encoding spastin gene, and polypeptides.
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -
 PS
 PS Claim 1; Fig 8; 76pp; English.
 CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSA) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with spastin
 CC protein or spastin gene expression or activity. The present sequence
 CC encodes mouse spastin as given in the present invention.
 XX
 SQ Sequence 11493 BP; 3599 A; 2281 C; 2387 G; 3226 T; 0 other;
 Query Match 7.0%; Score 35; DB 22; Length 11493;
 Best Local Similarity 100.0%; Pred. No. 1.9e-07;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 203 gattgacttattgatgagatgcactatccc 237
 Db 127 gattgacttattgatgagatgcactatccc 161
 RESULT 7
 AAH20130
 ID AAH20130 standard; DNA: 23 BP.
 XX
 AC AAH20130;
 XX
 DT 09-AUG-2001 (first entry)
 XX
 DE Human spastin ORF PCR primer SEQ ID NO:23.
 XX
 XX Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation;
 KW PCR primer; ss.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO200129266-A2.
 XX
 PD 26-APR-2001.
 XX
 PF 20-OCT-2000; 2000WO-US29130.
 XX

PR 20-OCT-1999; 9905-0160588.
 XX
 PA (UYMC-) UNIV MCGILL.
 PA (HOP-) HOPITAL SAINTE-JUSTINE.
 XX
 PI Hudson TJ, Engert J, Richter A;
 XX
 DR WPI; 2001-308494/32.
 XX
 PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -
 PS
 PS Claim 23; Fig 7; 76pp; English.
 CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSA) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with spastin
 CC protein or spastin gene expression or activity. AAH20122 to AAH20173
 CC represent specifically claimed primers which can be used in diagnostic
 CC methods from the present invention.
 XX
 SQ Sequence 23 BP; 7 A; 7 C; 3 G; 6 T; 0 other;
 Query Match 4.6%; Score 23; DB 22; Length 23;
 Best Local Similarity 100.0%; Pred. No. 0.17;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 392 gatgacatatacaacatccgct 414
 Db 1 gatgacatatacaacatccgct 23
 RESULT 8
 AAT35105
 ID AAT35105 standard; cDNA: 721 BP.
 XX
 AC AAT35105;
 XX
 DT 21-NOV-1996 (first entry)
 XX
 DE Down-regulated senescence clone, SBN31 proposed full length sequence.
 XX
 KW Senescence related gene; expression; inhibition; acceleration; tomato;
 KW lettuce; cabbage; banana; strawberry; wheat; maize; rice; rape; soybean;
 KW delayed senescence; yield; protein content; quality; tolerance;
 KW increased senescence; desiccant; ss.
 XX
 OS Lycopersicon esculentum.
 XX
 PN WO9507993-A1.
 XX
 PD 23-MAR-1995.
 XX

XX 13-SEP-1994; 94MO-GB01990.
PF 13-SEP-1993; 93GB-0018927.
XX (ZENE) ZENECA LTD.
PA
XX Drake CR, Farrell A, Grierson D, Hosken SE, John I;
PI Schuch WM, Smart CM, Thomas H;
XX WPI: 1995-111361/17.
DR
XX DNA constructs which modify expression of senescence-related genes -
PT useful to accelerate or inhibit senescence in crop plants
PS
XX Claim 10, Page 51; 78pp; English.
XX The sequences given in AAT35095-133 are senescence related genes and
CC fragments which were used in the construct of the invention which
CC modifies the expression of at least one senescence related gene.
CC Using these constructs senescence may be inhibited or accelerated
CC in plants including tomato, lettuce, cabbage, banana, strawberry,
CC wheat, maize, rice, rape or soybean. Delayed senescence may
CC indirectly prolong the life of the plant, increase yield, increase
CC protein content of fruits, improve quality of leafy vegetables.
CC Improve tolerance to disease, drought or other stress. Increased
CC senescence may more rapidly break down unwanted plant material and
CC so avoid the use of desiccants on crops. This sequence represents the
CC proposed full length sequence of the senescence down-regulating clone,
CC PSEN31, which is also known as 1M4. PSEN31 is a cDNA of approx. 0.9 kb
CC encoding a mRNA of approx. 1.0 kb. The mRNA encoded by PSEN31 is
CC expressed in green leaves of tomato plants, but at the onset of
CC senescence its expression is switched off. The PSEN31 sequence
CC exhibits 100% homology to the tomato cDNA clone TAS14 which is inducible
CC by salt stress and ABA in tomato seedlings. Southern analysis suggests
CC that there is one gene per haploid genome. This gene is specifically
CC reduced during tomato leaf senescence. PSEN31 is deposited as NCIMB
CC 40576.
XX
SQ Sequence 721 BP; 222 A; 115 C; 192 G; 192 T; 0 other;

Query Match 3.8%; Score 19; DB 16; Length 721;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 459 agataatgagagaatgcc 477
|||||
DB 303 agataatgagagaatgcc 321

RESULT 9
AAS83269
ID AAS83269 standard; cDNA; 1543 BP.
XX
AC AAS83269;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #19073.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001MO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI: 2001-639362/73.
DR
XX P-PADB; ABG19082.
DR
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics; forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID No 19073; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1543 BP; 489 A; 296 C; 374 G; 384 T; 0 other;

Query Match 3.8%; Score 19; DB 23; Length 1543;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 192 attttcagagagattgac 210
|||||
DB 1389 attttcagagagattgac 1407

RESULT 10
AAH72857
ID AAH72857 standard; cDNA; 1588 BP.
XX
AC AAH72857;
XX
DT 19-SEP-2001 (first entry)
XX
DE Human cervical cancer marker nucleic acid 4131.
XX
XX Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
KW
OS Homo sapiens.
XX
PN WO200142467-A2.
XX
PD 14-JUN-2001.
XX
XX 08-DEC-2000; 2000MO-US33312.
XX
XX 08-DEC-1999; 99US-0169681.
XX
XX 21-DEC-1999; 99US-0171350.
XX
PR 14-MAR-2000; 2000US-0189315.

```

PR 12-MAY-2000; 2000US-0203791.
PR 09-JUN-2000; 2000US-0210600.
PR 21-JUL-2000; 2000US-0220114.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Schlegel R, Deeds J, Berger A, Zhao X;
XX
DR WPI; 2001-375006/39.
XX
PT New isolated nucleic acid for diagnosing and treating cervical cancer
PT and for assessing and detecting compounds for treating the cancer -
XX
XX
PS Claim 1; Page 892-893; 1051pp; English.
XX
CC The invention relates to novel genes (AAH68727-AAH73383) associated with
CC cervical cancer with cytostatic activity. The nucleic acids and encoded
CC polypeptides are useful: to assess if a patient is afflicted with
CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition in a patient;
CC and to select and/or assess the efficacy of a compound or therapy for
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
CC useful for gene therapy.
XX
SQ Sequence 1588 BP; 511 A; 272 C; 356 G; 440 T; 9 other;

Query Match          3.8%; Score 19; DB 22; Length 1588;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 192 attttcagagatttcac 210
   |||||
DB 1365 attttcagagatttcac 1383

RESULT 11
AAV74557/C
ID AAV74557 standard; DNA; 3365 BP.
XX
AC AAV74557;
XX
DT 16-MAR-1999 (first entry)
XX
DE Staphylococcus aureus contig SEQ ID #246.
XX
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
XX
XX Staphylococcus aureus.
OS
FH Key Location/Qualifiers
FT m1sc-feature 421..480
   /tag= a
   /note= "these bases represent a line of missing text in
   the sequence listing in the specification. They
   are included to maintain the nucleotide numbering
   given in the specification for this DNA sequence"
FT m1sc-feature 2221..2280
   /tag= b
   /note= "these bases represent a line of missing text in
   the sequence listing in the specification. They
   are included to maintain the nucleotide numbering
   given in the specification for this DNA sequence"
XX
XX EP786519-A2.
XX
XX 30-JUL-1997.
XX
XX 07-JAN-1997; 97EP-0100117.
XX

```

```

PR 05-JAN-1996; 96US-0009861.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
XX
XX Rosen CA;
XX
DR WPI; 1997-374922/35.
XX
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
XX
XX
PS Claim 1; Page 1037-1039; 3271pp; English.
XX
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
XX
SQ Sequence 3365 BP; 1165 A; 536 C; 505 G; 1020 T; 139 other;

Query Match          3.8%; Score 19; DB 18; Length 3365;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 415 tatttaaaatataatcat 433
   |||||
DB 852 TATTAAAAAATATATTCAT 834

RESULT 12
ABL14512
ID ABL14512 standard; cDNA; 4057 BP.
XX
AC ABL14512;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 36018.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EM;
XX

```


XX WPI: 2001-6556860/75.
 DR P-PDB; ABB70409.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 38018; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB150511), expressed DNA
 CC sequences (AB161840-AB16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_ptc_sequences.
 CC
 XX Sequence 4057 BP; 1264 A; 799 C; 851 G; 1143 T; 0 other;
 SQ

Query Match 3.8%; Score 19; DB 23; Length 4057;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 183 ttatatcatcttcaga 201
 ||||||||||||||||
 DB 983 ttatatcatcttcaga 1001

RESULT 13
 ABL34145/c
 ID ABL34145 standard; DNA: 7076 BP.
 XX
 AC ABL34145;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human immune system associated gene SEQ ID NO: 2118.
 XX
 KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antihaemic; cytosolic; noctropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antineumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200200928-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP07537.
 XX
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI: 2002-130909/17.
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -

XX Claim 1; SEQ ID NO 2118; 32pp + Sequence Listing; German.
 PS
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 CC
 XX Sequence 7076 BP; 1994 A; 65 C; 1467 G; 3550 T; 0 other;
 SQ

Query Match 3.8%; Score 19; DB 24; Length 7076;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 179 aatcttatcatctttt 197
 ||||||||||||||||
 DB 4408 AATCTTATATACATTTT 4390

RESULT 14
 AAH12392
 ID AAH12392 standard; CDNA: 431 BP.
 XX
 AC AAH12392;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA clone (3'-primer) SEQ ID NO:9227.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Negai K, Otsuki T;
 XX
 DR WPI: 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 3; SEQ ID 9227; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification, where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences. AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

SQ Sequence 431 BP; 153 A; 84 C; 63 G; 126 T; 5 other;

Query Match 3.6%; Score 18; DB 22; Length 431;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 413 ctattataaaatatt 430
 ||||||||||||||||
 Db 1 ctattataaaatatt 18

RESULT 15

AAF55109/C
 ID AAF55109 standard; DNA: 513 BP.

AC AAF55109;

DT 15-MAY-2001 (first entry)

DE Nucleotide sequence of a BASB119 polypeptide.

KW BASB119; immune response; bacterial infection; genetic immunization;

KW otitis media; pneumonia; sinusitis; nosocomial infection;

KW invasive disease; hearing loss; fluid accumulation; antibacterial; ss.

OS Moraxella catarrhalis.

Key Location/Qualifiers

FT CDS 1..513
 FT /*tag= a
 FT /product= "BASB119"
 FT /note= "no termination codon given"

PN WO200109336-A1.

PD 08-FEB-2001.

PF 31-JUL-2000; 2000MO-EP07363.

PR 03-AUG-1999; 99GB-0018302.

PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

PI Thonnard J;

DR WPI; 2001-159873/16.

DR P-PSDB; AAB67488.

PT New BASB119 polypeptides and polynucleotides from Moraxella catarrhalis

PT strain ATCC 43617, useful as therapeutic agents or vaccines against

PT bacterial infections, e.g. otitis media or pneumonia -

PS Claim 13; Page 65; 83pp; English.

CC The present sequence encodes a BASB119 polypeptide of Moraxella

CC catarrhalis strain ATCC43617. BASB119 polypeptides and polynucleotides

CC are useful for generating an immune response in an animal.. The

CC polypeptides may also be used as prophylactic agents of bacterial

CC infections, particularly M. catarrhalis infections in mammals,
 CC especially humans. The polynucleotides are useful in therapy or
 CC prophylaxis, particularly genetic immunization against these infections
 CC or diseases. These diseases include otitis media in infants or
 CC children, pneumonia in elders, sinusitis, nosocomial infections and
 CC invasive diseases, chronic otitis media with hearing loss, fluid
 CC accumulation in the middle ear, infection of the upper respiratory
 CC tract, or inflammation of the middle ear. The polypeptides or
 CC polynucleotides may also be employed as research reagents and materials
 CC for discovering treatments of and diagnostics for diseases,
 CC particularly human diseases. In particular, the polypeptides or
 CC polynucleotides are useful in the discovery and development of the
 CC antibacterial compounds, or for diagnosing diseases, staging of the
 CC disease, determining the response of an infectious organism to drugs.

SQ Sequence 513 BP; 173 A; 78 C; 111 G; 151 T; 0 other;

Query Match 3.6%; Score 18; DB 22; Length 513;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 ttgatttgcaatggtat 121
 ||||||||||||||||
 Db 103 TTGATTGTCATGCTAT 86

Search completed: May 22, 2002, 08:35:42
 Job time: 6562 sec



THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 06:34:35 ; Search time 91.58 Seconds
(without alignments)
1341.087 Million cell updates/sec

Title: US-09-693-205-7_COPY_1_500

Perfect score: 500

Sequence: 1 atgattacaggaagacat.....gcagaattgtatcaaa 500

Scoring table: OLIGO_MNC
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/1/ina/6C_COMB.seq:*
6: /cgn2_6/prodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	3.6	3736	4	US-09-600-776-10
2	17	3.4	305	4	US-09-020-956-66
3	17	3.4	305	4	US-09-030-607-66
4	17	3.4	305	4	US-09-439-313-66
5	17	3.4	2442	1	US-08-390-162-5
6	17	3.4	2442	1	US-08-685-945B-5
7	17	3.4	8700	2	US-08-392-625-16
8	17	3.4	8700	2	US-08-466-961A-16
9	17	3.4	8700	2	US-08-645-193B-18
10	17	3.4	8835	3	US-08-884-324-14
11	17	3.4	28984	3	US-08-884-324-14
12	16	3.2	57	6	US-09-025-580-19
13	16	3.2	57	6	5432082-6
14	16	3.2	449	4	US-08-946-914-35
15	16	3.2	677	4	US-08-998-416-1092
16	16	3.2	685	4	US-09-327-357-100
17	16	3.2	821	4	US-08-998-416-541
18	16	3.2	872	4	US-08-998-416-487
19	16	3.2	1341	1	US-08-180-209B-26
20	16	3.2	1341	1	US-08-385-745-26
21	16	3.2	1341	4	US-08-485-388-26
22	16	3.2	1341	4	US-08-474-853-26
23	16	3.2	1341	5	PCT-US94-02629-26
24	16	3.2	1505	4	US-09-130-616-178
25	16	3.2	1773	4	US-09-130-616-177
26	16	3.2	1936	3	US-08-946-914-7
27	16	3.2	2156	1	US-08-321-356-1

28	16	3.2	2156	1	US-08-321-356-3	Sequence 3, App1
29	16	3.2	2372	4	US-09-130-616-174	Sequence 174, App
30	16	3.2	3450	2	US-08-378-817A-9	Sequence 9, App1
31	16	3.2	3592	3	US-08-714-918-63	Sequence 63, App1
32	16	3.2	3592	4	US-09-265-315-63	Sequence 63, App1
33	16	3.2	3592	4	US-09-265-315-63	Sequence 63, App1
34	16	3.2	3592	4	US-09-265-315-63	Sequence 63, App1
35	16	3.2	3906	2	US-08-469-537A-102	Sequence 102, App
36	16	3.2	4060	1	US-08-164-292B-1	Sequence 1, App1
37	16	3.2	4060	1	US-08-164-292B-3	Sequence 3, App1
38	16	3.2	4060	1	US-08-164-292B-5	Sequence 5, App1
39	16	3.2	4060	1	US-08-164-292B-7	Sequence 7, App1
40	16	3.2	4060	3	US-08-845-623-1	Sequence 1, App1
41	16	3.2	4060	3	US-08-845-623-3	Sequence 3, App1
42	16	3.2	4060	3	US-08-845-623-5	Sequence 5, App1
43	16	3.2	4060	3	US-08-845-623-7	Sequence 7, App1
44	16	3.2	4060	3	US-08-815-927-1	Sequence 1, App1
45	16	3.2	4060	3	US-08-815-927-3	Sequence 3, App1

ALIGNMENTS

RESULT 1
US-09-600-776-10
; Sequence 10, Application US/09600776
; Patent No. 6326168
; GENERAL INFORMATION:
; APPLICANT: Yamanouchi Pharmaceutical Co., Ltd.
; TITLE OF INVENTION: A novel potassium channel protein
; FILE REFERENCE: Y9903-PCT
; CURRENT APPLICATION NUMBER: US/09/600,776
; CURRENT FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: JP P1998-011434
; PRIOR FILING DATE: 1998-01-23
; PRIOR APPLICATION NUMBER: JP P1998-346198
; PRIOR FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 10
; LENGTH: 3736
; TYPE: DNA
; ORGANISM: Rattus sp.
US-09-600-776-10

Query Match 3.6%; Score 18; DB 4; Length 3736;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 321 ctgaagcagcttcacg 338
|||
Db 3148 ctgaagcagcttcacg 3165

RESULT 2
US-09-020-956-66/c
; Sequence 66, Application US/09020956
; Patent No. 6261562
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
; NUMBER OF SEQUENCES: 178
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/020,956
FILING DATE: 09-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.427C2
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 305 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-09-020-956-66

Query Match 3.4%; Score 17; DB 4; Length 305;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 413 ctattataaaatatat 429
|||||
DB 259 CTTATTAAAAATATAT 243

RESULT 3
US-09-030-607-66/C
Sequence 66, Application US/09030607
Patent No. 6262245
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
NUMBER OF SEQUENCES: 224
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/030,607
FILING DATE: 25-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.427C3
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 305 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-09-030-607-66

Query Match 3.4%; Score 17; DB 4; Length 305;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 413 ctattataaaatatat 429
|||||
DB 259 CTTATTAAAAATATAT 243

RESULT 4
US-09-439-313-66/C
Sequence 66, Application US/09439313
Patent No. 6329505
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan Louise
APPLICANT: Jiang Yuqun
APPLICANT: Reed, Steven G.
APPLICANT: Kalos, Michael
APPLICANT: Fanger, Gary
APPLICANT: Retter, Mark
APPLICANT: Solk, John
APPLICANT: Day, Craig
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C9
CURRENT APPLICATION NUMBER: US/09/439,313
CURRENT FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 575
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 66
LENGTH: 305
TYPE: DNA
ORGANISM: Homo sapien
US-09-439-313-66

Query Match 3.4%; Score 17; DB 4; Length 305;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 413 ctattataaaatatat 429
|||||
DB 259 CTTATTAAAAATATAT 243

RESULT 5
US-08-390-162-5/C
Sequence 5, Application US/08390162
Patent No. 5576192
GENERAL INFORMATION:
APPLICANT: Ichikawa, Atsushi
APPLICANT: Narumiya, Shuh
TITLE OF INVENTION: Prostaglandin E Receptors, Their DNA and
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESS: Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/390,162
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/024179
FILING DATE: 23-FEB-1993
APPLICATION NUMBER: JP 036580-1992
FILING DATE: 24-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 064889-1992
FILING DATE: 23-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Fordis, Jean B.
REGISTRATION NUMBER: 32,984
REFERENCE/DOCKET NUMBER: 04221-0020-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2442 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-390-162-5

Query Match 3.4%; Score 17; DB 1; Length 2442;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Caps 0;

OY 185 tataacatttttcaga 201
|||||
Db 1674 TATATACATTTTTCAGA 1658

RESULT 6
US-08-685-945B-5/C
Sequence 5, Application US/08685945B
Patent No. 5804415
GENERAL INFORMATION:
APPLICANT: Ichikawa, Atsushi
APPLICANT: Natumaiya, Shuh
TITLE OF INVENTION: Prostaglandin E Receptors, Their DNA and
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/685,945B
FILING DATE: 22-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/024179
FILING DATE: 23-FEB-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 036580-1992

FILING DATE: 24-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 064889-1992
FILING DATE: 23-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Fordis, Jean B.
REGISTRATION NUMBER: 32,984
REFERENCE/DOCKET NUMBER: 04221-0020-02000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2442 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-685-945B-5

Query Match 3.4%; Score 17; DB 1; Length 2442;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Caps 0;

OY 185 tataacatttttcaga 201
|||||
Db 1674 TATATACATTTTTCAGA 1658

RESULT 7
US-08-392-625-16/C
Sequence 16, Application US/08392625
Patent No. 5837485
GENERAL INFORMATION:
APPLICANT: Entian, Karl-Dieter
APPLICANT: G tz, Friedrich
APPLICANT: Schnell, No. 5837485bert
APPLICANT: Augustin, Johannes
APPLICANT: Engelke, Germar
APPLICANT: Rosensteel, Ralf
APPLICANT: Kaletta, Cortina
APPLICANT: Klein, Cora
APPLICANT: Wieland, Bernd
APPLICANT: Kupke, Thomas
APPLICANT: Jung, G nther
APPLICANT: Kellner, Roland
TITLE OF INVENTION: Biosynthetic Process For The Preparation
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/392,625
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/876,791
FILING DATE: 30-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 0652.0980002

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 8700 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-392-625-16

Query Match 3.4%; Score 17; DB 2; Length 8700;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 415 tattaaaaatatattc 431
|||||
DB 574 TATTAAAAATATTTC 558

RESULT 8
US-08-466-961A-16/c
Sequence 16, Application US/08466961A
Patent No. 5843709
GENERAL INFORMATION:
APPLICANT: Entlan, Karl-Dieter
APPLICANT: G t z, Friedrich
APPLICANT: Schnell, No. 5843709bert
APPLICANT: Augustin, Johannes
APPLICANT: Engelke, Gernar
APPLICANT: Rosensteln, Ralf
APPLICANT: Kaletta, Corlina
APPLICANT: Klein, Cora
APPLICANT: Wieland, Bernd
APPLICANT: Kupke, Thomas
APPLICANT: Jung, G nther
TITLE OF INVENTION: Biosynthetic Process for the Preparation of
TITLE OF INVENTION: Chemical Compounds
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, NW
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,961A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: US 08/392,625
FILING DATE: 22-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/876,791
FILING DATE: 30-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/784,234
FILING DATE: 31-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 0652.0980004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 8700 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-466-961A-16

Query Match 3.4%; Score 17; DB 2; Length 8700;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 415 tattaaaaatatattc 431
|||||
DB 574 TATTAAAAATATTTC 558

RESULT 9
US-08-645-193B-18/c
Sequence 18, Application US/08645193B
Patent No. 5962253
GENERAL INFORMATION:
APPLICANT: Kupke, Thomas
APPLICANT: Gotz, Friedrich
APPLICANT: Kempler, Christoph
APPLICANT: Jung, Gunther
TITLE OF INVENTION: Oxidative Decarboxylation of Peptides
TITLE OF INVENTION: Catalyzed by Flavoprotein Epld
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/645,193B
FILING DATE: 13-MAY-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 0652.1540000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 8700 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: cDNA
US-08-645-193B-18

Query Match 3.4%; Score 17; DB 2; Length 8700;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 415 tattaaaaatatattc 431
|||||
DB 574 TATTAAAAATATTTC 558

RESULT 10

US-08-884-324-10
; Sequence 10, Application US/08884324
; Patent No. 6060283
; GENERAL INFORMATION:
; APPLICANT: Takano, OKURA
; APPLICANT: Kakui TORIGOE
; APPLICANT: Masashi KURIMOTO
; TITLE OF INVENTION: GENOMIC DNA ENCODING A POLYPEPTIDE CAPABLE
; OF INDUCING THE PRODUCTION OF INTERFERON-
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/884,324
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 185,305/96
; FILING DATE: 27-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: OKURA-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8835 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: human
; TISSUE TYPE: placenta
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 1..8835
; IDENTIFICATION METHOD: E
; US-08-884-324-10

Query Match 3.4%; Score 17; DB 3; Length 8835;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 ttctgacctgacagaga 102
|||||
Db 3604 TTCTGCGCTGCGACAGA 3620

RESULT 11
US-08-884-324-14
; Sequence 14, Application US/08884324
; Patent No. 6060283
; GENERAL INFORMATION:
; APPLICANT: Takano, OKURA
; APPLICANT: Kakui TORIGOE
; APPLICANT: Masashi KURIMOTO
; TITLE OF INVENTION: GENOMIC DNA ENCODING A POLYPEPTIDE CAPABLE
; OF INDUCING THE PRODUCTION OF INTERFERON-
; NUMBER OF SEQUENCES: 35

CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/884,324
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 185,305/96
FILING DATE: 27-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: OKURA-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 28994 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: human
TISSUE TYPE: placenta
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..15606
IDENTIFICATION METHOD: E
NAME/KEY: leader peptide
LOCATION: 15607..15685
IDENTIFICATION METHOD: S
NAME/KEY: Intron
LOCATION: 15686..17056
IDENTIFICATION METHOD: E
NAME/KEY: leader peptide
LOCATION: 17057..17068
IDENTIFICATION METHOD: S
NAME/KEY: Intron
LOCATION: 17069..20451
IDENTIFICATION METHOD: E
NAME/KEY: leader peptide
LOCATION: 20452..20468
IDENTIFICATION METHOD: S
NAME/KEY: mat peptide
LOCATION: 20469..20586
IDENTIFICATION METHOD: S
NAME/KEY: Intron
LOCATION: 20587..21920
IDENTIFICATION METHOD: E
NAME/KEY: mat peptide
LOCATION: 21921..22054
IDENTIFICATION METHOD: S
NAME/KEY: Intron
LOCATION: 22055..26827
IDENTIFICATION METHOD: E
NAME/KEY: mat peptide
LOCATION: 26828..27046
IDENTIFICATION METHOD: S
NAME/KEY: 3'UTR
LOCATION: 27047..28994
IDENTIFICATION METHOD: E

US-08-884-324-14

Query Match 3.4%; Score 17; DB 3; Length 28994;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 ttctgcctgcgcagaga 102
|||||
DB 10367 TTCTGGCTGCGAGAGA 10383

RESULT 12

US-09-025-580-19/c
; Sequence 19, Application US/09025580
; Patent No. 6162613
; GENERAL INFORMATION:
; APPLICANT: Su, Michael Shin-San
; APPLICANT: Fox, Ted
; APPLICANT: Wilson, Keith Phillip
; APPLICANT: Germann, Ursula A.
; TITLE OF INVENTION: Methods For Designing Inhibitors of
; TITLE OF INVENTION: Serine/Threonine Kinases and Tyrosine Kinase
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/025,580
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: VPI 97-104
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 37 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "oligonucleotide"
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-09-025-580-19

Query Match 3.2%; Score 16; DB 4; Length 37;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 caaggaagtaatgaat 82
|||||
DB 35 CAAGGAAGTAATGAAT 20

RESULT 13
5432082-6/c
; Patent No. 5432082
; APPLICANT: GALEBOTTI, CESIRA, PALIA, EMANUELA, RAUGEI, GIOVANNI;

; BENSI, GIULIANO; MELLI, MARIA L.
; TITLE OF INVENTION: EXPRESSION AND SECRETION VECTOR IN YEASTS,
; USEFUL FOR PREPARING HETEROLOGOUS PROTEINS

; NUMBER OF SEQUENCES: 6
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/69,455
; FILING DATE: 01-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 654,069; 69,329
; FILING DATE: 11-FEB-1991; 02-JUL-1987
; APPLICATION NUMBER: 69,329
; FILING DATE: 02-JUL-1987
; SEQ ID NO: 6:
; LENGTH: 57

5432082-6

Query Match 3.2%; Score 16; DB 6; Length 57;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 418 taaaaaatatcat 433
|||||
DB 16 TAAAAATATATTCAT 1

RESULT 14

US-08-946-914-35/c
; Sequence 35, Application US/08946914
; Patent No. 6027916
; GENERAL INFORMATION:
; APPLICANT: Ni, Jian
; APPLICANT: Gentz, Reiner L.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.
; STREET: 1100 New York Ave., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/946,914
; FILING DATE: Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/028,093
; FILING DATE: 09-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488, 0560001/EKS/GW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 449 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-946-914-35

Query Match 3.2%; Score 16; DB 3; Length 449;

Best Local Similarity 100.0%; Pred. No. 78;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 catttagacgatgaa 319
|||||
Db 54 CATTTAGACGATGAA 39

RESULT 15

US-08-998-416-1092/C
; Sequence 1092, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Reblschung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSTYPI
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264tn Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/GC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 1092:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 677 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1666UP
; US-08-998-416-1092

Query Match 3.2%; Score 16; DB 4; Length 677;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 415 tatataaataatatt 430
|||||
Db 442 TATTAATAAATAATATT 427

Search completed: May 22, 2002, 08:29:03
Job time: 6868 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 04:30:30 ; Search time 2968.03 Seconds
(without alignments)
2273.723 Million cell updates/sec

Title: US-09-693-205-7_COPY_1_500
Perfect score: 500
Sequence: 1 atgatttcaggaagaccat.....gcagaattcgttaacaa 500

Scoring table:
OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 13736207 segs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST*
1: em_estha:*
2: em_esthum:*
3: em_estlin:*
4: em_estlun:*
5: em_estlov:*
6: em_estlpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_estl2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pin:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	176	35.2	241	9	BE008891
C 2	57	11.4	266	10	BF326199
C 3	42	8.4	746	12	BH025017
C 4	42	8.4	796	12	BH126217
C 5	35	7.0	558	12	AZ649875
C 6	22	4.4	202	9	BB561400
C 7	22	4.4	440	12	AZ066617
C 8	22	4.4	496	12	AZ644393
C 9	22	4.4	717	12	BH110525
C 10	20	4.0	170	10	BF397419
C 11	20	4.0	347	10	BF703038
C 12	20	4.0	630	10	BG599717
C 13	20	4.0	630	10	BG889484
C 14	20	4.0	686	9	AW018508
C 15	20	4.0	728	12	BH348265
C 16	20	4.0	755	10	BG593815
C 17	20	4.0	968	12	CNS06FVJ

18	19	3.8	189	9	BE152099
C 19	19	3.8	193	9	AA481759
C 20	19	3.8	220	9	AI275781
C 21	19	3.8	252	9	AL642337
C 22	19	3.8	271	10	R96076
C 23	19	3.8	280	9	AI823324
C 24	19	3.8	287	9	AI932736
C 25	19	3.8	282	9	AV295506
C 26	19	3.8	303	9	AV304343
C 27	19	3.8	323	10	H86539
C 28	19	3.8	328	10	BG941630
C 29	19	3.8	331	10	BG942635
C 30	19	3.8	342	9	AI468151
C 31	19	3.8	342	10	BG133892
C 32	19	3.8	351	9	AW262504
C 33	19	3.8	357	9	AI282243
C 34	19	3.8	358	9	AI042486
C 35	19	3.8	358	9	AI753035
C 36	19	3.8	360	9	AV194703
C 37	19	3.8	368	10	BG941160
C 38	19	3.8	372	10	N67665
C 39	19	3.8	375	10	N24075
C 40	19	3.8	377	9	AI025218
C 41	19	3.8	379	9	AI095227
C 42	19	3.8	379	10	W32532
C 43	19	3.8	381	9	AW204595
C 44	19	3.8	384	12	AQ129047
C 45	19	3.8	400	12	BH316467

ALIGNMENTS

RESULT 1
LOCUS BE008891
DEFINITION CM4-BN0161-040400-132-d08 BN0161 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE008891
VERSION BE008891.1 GI:8269124
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Nagai,M.A., da Silva,M.Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-CM4-BN0161-040400-132-d08&ts=2000-04-04&tl=1)

Seq primer: puc 18 forward
High quality sequence start: 6
High quality sequence stop: 241.
Location/Qualifiers
1. .241
/organism="Homo sapiens"

FEATURES
source

/db_xref="taxon:9606"
/clone_lib="BN0161"
/dev_stage="Adult"
/note="Organ: breast normal; Vector: puc18; Site: 1: SmaI; Site: 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 73 a 40 c 47 g 81 t
ORIGIN

Query Match 35.2%; Score 176; DB 9; Length 241;
Best Local Similarity 100.0%; Pred. No. 3.3e-81;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 275 ctcattagactcagatcttcattagcattagcagatgaatcgaagcagcgt 334
|||||
Db 241 CTCATTAGACTCAGATTCATGCTTAGTCATTAGACATGATCGAAGCAGACTT 182
|||||

QY 335 ccgaatttttagcagacatgtacaaaactggagggtttgtccttaaaattagat 394
|||||
Db 181 CCGAAATTTTACGACATGTGACAAAACCTGGAGGTTTCTTAAAAAATTAGAT 122
|||||

QY 395 gcatctatacaacatcgctattaaataatattcattcacattaccagttgc 450
|||||
Db 121 GCATCTATACACATCCGCTTATTAAAAATATATTCATTCCACATTACCAAGTGC 66
|||||

RESULT 2
BF326199 266 bp mRNA linear EST 22-NOV-2000
LOCUS MRO-AN0083-160900-003-d10 AN0083 Homo sapiens cDNA, mRNA, sequence.
DEFINITION BF326199
ACCESSION BF326199
VERSION BF326199.1 GI:11296947
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 266)
Dias Neto, E., Garcia Correia, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, M. Jr., Zago, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Baia, G. S., Simpson, D. H., Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL MEDLINE
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MRO-AN0083-160900-003-d10&ts=2000-09-16&td=1)
Seq primer: puc 18 forward
High quality sequence stop: 11.
Location/Qualifiers
1. 266
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="AN0083"
/dev_stage="Adult"

FEATURES
source

/note="Organ: amnion normal; Vector: puc18; Site: 1: SmaI; Site: 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 71 a 67 c 52 g 76 t
ORIGIN

Query Match 11.4%; Score 57; DB 10; Length 266;
Best Local Similarity 100.0%; Pred. No. 7.8e-19;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 aggaagccatctactagctgcagctctcaatccgaagcgtttgcagctcttat 66
|||||
Db 210 AGGAAGCAGCATGTACGCTGCGAGCTTAAATCCGAGAGATTGCGAGCTTAT 266
|||||

RESULT 3
BH025017/c 746 bp DNA linear GSS 17-JUL-2001
LOCUS RPCI-24-318E15.TJ RPCI-24 Mus musculus genomic clone RPCI-24-318E15
DEFINITION
ACCESSION BH025017
VERSION BH025017.1 GI:14788481
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 746)
Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.
Zhao, S., Nieman, M., Malek, J., Shatsman, S., Akiret, B., Levins, M., Tsegaye, G., Geer, R., Krol, M., Shvartsbeyn, A., Gebregorgis, E., Russell, D., de Jong, P. and Fraser, C. M.
Mouse BAC End sequences from Library RPCI-24
Unpublished (1999)
Other GSSs: RPCI-24-318E15.TVB
Contact: Shanying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.choi.org/bacpac/orderinfoframe.htm). BAC end page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
Plate: 318 row: E column: 15
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1. 746
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone=RPCI-24-318E15
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pPRBAC1; Site: 1: BamHI; Site: 2: BamHI; The library was cloned in the pPRBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

FEATURES
source

BASE COUNT 229 a 142 c 137 g 238 t
ORIGIN

Query Match 8.4%; Score 42; DB 12; Length 746;

Best Local Similarity 100.0%; Pred. No. 6.1e-11;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 tacaggaagaccatgtactcagctcagcttctaataccaga 48
|||||
Db 591 TACAGGAAGACATGTACTCAGCTGACGCTTCTTAATCCAGA 550

RESULT 4

BH126217/c 796 bp DNA linear GSS 19-JUL-2001
LOCUS
DEFINITION
RPCI-24-318C15.TJ RPCI-24 Mus musculus genomic clone RPCI-24-318C15
, DNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
BHI26217.1 GI:14969729
GSS.
house mouse.
Mus musculus

REFERENCE
AUTHORS
1 (bases 1 to 796)
Zhao,S., Nieman,W., Malek,J., Shatsman,S., Akiret,B., Levins,M.,
Tsegaye,G., Geer,R., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)

TITLE
JOURNAL
COMMENT
Other-GSSs: RPCI-24-318C15.TVB
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPC
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
page: http://www.choi.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 318 row: C column: 15
Seq primer: SP6
Class: BAC ends.

FEATURES
source
Location/Qualifiers
1..796

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-318C15"
/clone_1lb="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTRABAC1; Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTRABAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."

BASE COUNT 244 a 151 c 144 g 257 t
ORIGIN

Query Match 8.4%; Score 42; DB 12; Length 796;
Best Local Similarity 100.0%; Pred. No. 6.1e-11;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 tacaggaagaccatgtactcagctcagcttctaataccaga 48
|||||
Db 614 TACAGGAAGACATGTACTCAGCTGACGCTTCTTAATCCAGA 573

RESULT 5
LOCUS A2649875 558 bp DNA linear GSS 14-DEC-2000
DEFINITION IM0519B14R Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC1M0519B14 R, DNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
A2649875.1 GI:11783794
GSS.
house mouse.
Mus musculus

REFERENCE
AUTHORS
1 (bases 1 to 558)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rellily,
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)

TITLE
JOURNAL
COMMENT
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0519 row: B column: 14
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 558.

FEATURES
source
Location/Qualifiers
1..558

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0519B14"
/clone_1lb="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(<http://www.jax.org/resources/documents/dnares/>). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (g114732114[9b]AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 200 a 99 c 96 g 163 t
ORIGIN

Query Match 7.0%; Score 35; DB 12; Length 558;
Best Local Similarity 100.0%; Pred. No. 2.8e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 203 gatttgacttatttgatgagatccactatccc 237
|||||
Db 51 GATTTGACTTATTGTGATGAGATCCACTATGCC 85

RESULT 6
LOCUS BB561400 202 bp mRNA linear EST 01-AUG-2000
DEFINITION BB561400

DEFINITION	BBS61400 RIKEN full-length enriched, 10 days neonate olfactory brain Mus musculus cDNA clone E530213F07 3', mRNA sequence.
ACCESSION	BBS61400
VERSION	BBS61400.1
KEYWORDS	EST.
SOURCE	house mouse. Mus musculus
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 302)
AUTHORS	Kono,H., Aikawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Harai,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,U., Ishikawa,T., Itoh,M., Izawa,M., Kacoti,K., Kagawa,I., Kai,C., Kawajiri,K., Kikuchi,N., Kiyosawa,H., Koijima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M., Matuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sobge,Y., Suehara,Y., Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toya,T., Tsunoda,Y., Watanuki,A., Watanabe,S., Yamamura,T., Yamanaka,I., Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE	RIKEN Mouse ESTs (Kono,H., et al.)
JOURNAL	Unpublished (2000)
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suicho-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gs.c.riken.go.jp, URL:http://genome.gsc.riken.go.jp/ Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoaka,S., Sasakura,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Thermotabilization and thermoactivation of the labile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998) Itoh,M., Katsunami,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y. Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999) Carninci,P. and Hayashizaki,Y. High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999) Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES	Location/Qualifiers
SOURCE	1..202 /organism="Mus musculus" /db_xref="taxon:10090" /clone="E530213F07" /clone_lib="RIKEN full-length enriched, 10 days neonate olfactory brain" /tissue_type="olfactory brain" /dev_stage="10 days neonate" /lab_host="DH10B" /note="Site.1: SalI. Site.2: BamHI. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGGAGAGAGCGCCGGCTACTGCAGTTTTCCTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGGAGAGAGATTCGAGTTAATTAAATTAATCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified phluescript KS(+) after bulk excision from lambda FLC I."

BASE COUNT	62 a	53 c	29 g	57 t	1 others
ORIGIN					
Query Match	4.4%; Score 22; DB 9; Length 202;				
Best Local Similarity	100.0%; Pred. No. 1.7;				
Matches	22; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
OY	453	ttttgcagataatgagaagat 474			
DB	90	TTTTTCAGATATATGAGAAGAT 111			
RESULT 7					
AZ066617/c	AZ066617	440 bp	DNA	linear	GSS 30-MAR-2000
LOCUS					
DEFINITION	RPCI-23-433L3.TV RPCI-23 Mus musculus genomic clone RPCI-23-433L3,				
ACCESSION	DNA sequence.				
VERSION	AZ066617				
KEYWORDS	AZ066617.1	GI:7357869			
SOURCE	GSS.				
ORGANISM	house mouse.				
	Mus musculus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 440)				
AUTHORS	Zhao,S., Niemman,W., Feldbljum,T., Malek,J., Shatsman,S., Akintet ,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P. and Fraser,C.M.				
TITLE	Mouse BAC End Sequences from Library RPCI-23				
JOURNAL	Unpublished (1999)				
COMMENT	Other_GSS: RPCI-23-433L3.TJ Contact: Shaying Zhao Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: szhao@tigr.org Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html Plate: 433 row: L column: 3 Seq primer: T7 Class: BAC ends.				
FEATURES	Location/Qualifiers				
SOURCE	1..440				
	/organism="Mus musculus"				
	/strain="C57BL/6J"				
	/db_xref="taxon:10090"				
	/clone="RPCI-23-433L3"				
	/clone_lib="RPCI-23"				
	/sex="Female"				
	/lab_host="DH10B"				
	/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."				
BASE COUNT	120 a	74 c	73 g	173 t	
ORIGIN					
Query Match	4.4%; Score 22; DB 12; Length 440;				
Best Local Similarity	100.0%; Pred. No. 1.8;				
Matches	22; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
OY	453	ttttgcagataatgagaagat 474			

Db 411 TTTTGCAGATATGAGAGAT 390

RESULT 8

AZ644393 496 bp DNA linear GSS 14-DEC-2000
LOCUS 1M0508M18F Mouse 10kb plasmid UGCLM1 library Mus musculus genomic
DEFINITION clone UGCLM0508M18 F, DNA sequence.

ACCESSION AZ644393
VERSION AZ644393.1 GI:11772878

KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 496)
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A., and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SIC, UT 84112, USA

Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunne@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0508 row: M column: 18
Seq primer: CGTTGTAACGACGCGCCACT
Class: plasmid ends
High quality sequence stop: 496.
Location/Qualifiers

FEATURES

source

1. 496
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGCLM0508M18"
/clone_11b="Mouse 10kb plasmid UGCLM1 library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F."
/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g114732114191AF129072.1), a copy-number inducible derivative of plasmid RL. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 181 a 94 c 85 g 136 t
ORIGIN

Query Match 4.4%; Score 22; DB 12; Length 496;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 453 ttgtcagataatgagagaat 474
Db 162 TTTTGCAGATATGAGAGAT 183

RESULT 9

BH110525/c 717 bp DNA linear GSS 19-JUL-2001
LOCUS RPCI-24-340G10.TV RPCI-24 Mus musculus genomic clone RPCI-24-340G10
DEFINITION , DNA sequence.

ACCESSION BH110525
VERSION BH110525.1 GI:14944731

KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 717)
AUTHORS Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akintet, B., Levins, M., Tsegaye, G., Geer, K., Krol, M., Shwartsbeyn, A., Gebregeorgis, E., Russell, D., de Jong, P., and Fraser, C.M.

TITLE Mouse BAC End Sequences from Library RPCI-24
JOURNAL Unpublished (1999)
COMMENT Other GSSs: RPCI-24-340G10.TV

Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 340 row: G column: 10
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers

FEATURES

source

1. 717
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-340G10"
/clone_11b="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTRABAC1; Site 1: BamHI; Site 2: BamHI; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTRABAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 206 a 125 c 133 g 253 t
ORIGIN

Query Match 4.4%; Score 22; DB 12; Length 717;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 453 ttgtcagataatgagagaat 474
Db 539 TTTTGCAGATATGAGAGAT 518

RESULT 10

BF397419/c 170 bp mRNA linear EST 27-NOV-2000
LOCUS UI-R-BS2-bel-b-03-0-UI.s1 UI-R-BS2 Rattus norvegicus cDNA clone
DEFINITION UI-R-BS2-bel-b-03-0-UI 3', mRNA sequence.

ACCESSION BF397419
VERSION BF397419.1 GI:11382402

KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 170)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msosares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized embryo at 13 dpc library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=yes.

FEATURES
source
Location/Qualifiers
1..170
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UT-R-B52-bel-b-03-0-UI"
/clone_lib="UT-R-B52"
/dev_stage="Embryonic 13 dpc"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site.1: Not I; Site.2: Eco RI; The UT-R-B52 library is a subtracted library derived from 13 dpc whole embryo tissue. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LIB="UT-R-B52"
TAG_TISSUE="embryo at 13 dpc"
TAG_SEQ="ATATCC"
BASE COUNT 43 a 29 c 31 g 67 t
ORIGIN

Query Match 4.0%; Score 20; DB 10; Length 170;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 453 ttgtgcagataatgaggaag 472
|||||
Db 134 TTTTGCAGATATGAGAG 115

RESULT 11
LOCUS BF703038 347 bp mRNA linear EST 22-DEC-2000
DEFINITION MI-P-E5-abn-e-02-1-UM.s1 MI-P-E5 Sus scrofa cDNA clone
VERSION BF703038
KEYWORDS BF703038.1 GI:11988446
SOURCE EST.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 347)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Tuggle CK
Molecular Genetics Laboratory, Department of Animal Science
Iowa State University
201 Kildee Hall, Ames, IA 50011-3150, USA
Tel: 5152944252
Fax: 5152942401
Email: cktuggle@iastate.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the non-normalized embryo at gestational day 12 library cDNA Library Preparation: RJ Woods, JA Green, RS Prather SI42 Animal Science Research Center, Department of Animal Science, University of Missouri-Columbia, 65211 Clone distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 1-38,
>AT_rich#low_complexity
Seq primer: M13 Forward
POLYA=yes.

FEATURES
source
Location/Qualifiers
1..347
/organism="Sus scrofa"
/strain="crossbred"
/db_xref="taxon:9823"
/clone="MI-P-E5-abn-e-02-1-UM"
/clone_lib="MI-P-E5"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site.1: Not I; Site.2: EcoRI; The MI-P-E5 library is derived from embryo at gestational day 12. For a detailed description of the library from which this clone was derived, please visit our web site at <http://pigest.genome.iastate.edu/>.
TAG_LIB="MI-P-E5"
TAG_TISSUE="embryo at gestational day 12"
TAG_SEQ="GTGAGA"
BASE COUNT 129 a 80 c 67 g 71 t
ORIGIN

Query Match 4.0%; Score 20; DB 10; Length 347;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 176 aaaaatcttatatacatct 195
|||||
Db 12 AAAAATCTTATATACATTT 31

RESULT 12
LOCUS BG599717 630 bp mRNA linear EST 12-APR-2001
DEFINITION EST504612 cSTS Solanum tuberosum cDNA clone cSTS26K9 5' sequence,
VERSION BG599717
KEYWORDS BG599717.1 GI:13616853
SOURCE EST.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
1 (bases 1 to 630)
van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Chienlingo,A.,

sites of the pSPOR1 vector (BRU). Library was constructed by Matthew Clark (Lehrach Lab; ICRF, London and Max Planck Institut fuer Molekulare Genetik, Berlin). cDNAs for EST analysis were selected following oligonucleotide hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss), adult liver or embryonic shield stage (5.6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were sequenced to assess clustering parameters or single clones were sequenced additional times to assess quality control."

BASE COUNT 200 a 148 c 173 g 162 t 3 others
ORIGIN

Query Match 4.0%; Score 20; DB 9; Length 686;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 310 agacgatgaatctgaagcac 329
|||||
Db 574 AGACGATGAATCTGAAGCAC 593

RESULT 15
BH348265/C 728 bp DNA linear GSS 03-DEC-2001
LOCUS CH230-120F14.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
DEFINITION CH230-120F14, DNA sequence.
ACCESSION BH348265
VERSION BH348265.1 GI:17278999
KEYWORDS GSS.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 728)
AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
Jong,P. and Fraser,C.M.
TITLE Rat BAC End Sequences from Library CHORI-230 EcoRI segment
JOURNAL Unpublished (1999)
COMMENT other_GSSs: CH230-120F14.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pjejong@email.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or_ering_information.htm). BAC end
page: http://www.tigr.org/tdb/bac-ends/rat/bac_end_intro.html
Plate: 120 row: F column: 14
Seq primer: SP6
Class: BAC ends.

FEATURES
source Location/Qualifiers

1..728
/organism="Rattus norvegicus"
/strain="BN/SSNHSd/MCW"
/db_xref="taxon:10116"
/clone="CH230-120F14"
/clone_lib="CHORI-230 Segment 1"
/sex="Female"
/cell_type="Brain"
/note="Vector: PTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
CHORI-230 Rat (BN/SSNHSd/MCW) BAC library produced by

Pieter de Jong"
BASE COUNT 273 a 103 c 145 g 207 t
ORIGIN

Query Match 4.0%; Score 20; DB 12; Length 728;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 415 tattaaataatattcatt 434
|||||
Db 115 TATTAAAAATATATTCATT 96

Search completed: May 22, 2002, 07:30:25
Job time: 10795 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 08:35:42 ; Search time 373 Seconds
(without alignments)
6448.786 Million cell updates/sec

Title: US-09-693-205-7_COPY_5300_6700

Perfect score: 1401

Sequence: 1 gtacgcagtaaacactagagc.....aaagacaccttaagaagt 1401

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1736436 segs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

N_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB	ID	Description
1	1401	100.0	12792	22	AAH20176	Human mutated spastin
2	1295	92.4	12793	22	AAH20174	Human spastin nucleotide
3	1295	92.4	12793	22	AAH20178	Human mutated spastin
4	1295	92.4	12793	22	AAH20179	Human mutated spastin
5	1295	92.4	12793	22	AAH20182	Human mutated spastin
6	53	3.8	11493	22	AAH20175	Mouse spastin nucleotide
7	24	1.7	24	22	AAH20151	Human spastin ORF
8	21	1.5	21	22	AAH20149	Human spastin ORF
9	21	1.5	21	22	AAH20150	Human spastin ORF

10	21	1.5	21	22 AAH20152	Human spastin ORF
11	20	1.4	20	22 AAH20154	Human spastin ORF
12	20	1.4	20	22 AB125800	Drosophila melanog
13	19	1.4	19	22 AB166282	Novel human polynu
14	19	1.4	19	22 AA583961	DNA encoding novel
15	19	1.4	19	22 AA586530	C glutamicum codin
16	19	1.4	19	22 AA583963	DNA encoding novel
17	19	1.4	19	22 AA583963	Human polynucleoti
18	19	1.4	19	22 AA158872	Human polynucleoti
19	19	1.4	19	22 AA47414	Sequence encoding
20	19	1.4	19	22 AAH73238	Human cervical can
21	19	1.4	19	22 AA536912	Human cardiovascular
22	19	1.4	19	22 AA536912	C glutamicum codin
23	19	1.4	19	22 AA536912	Human differential
24	18	1.3	18	22 AAH81581	Human polynucleoti
25	18	1.3	18	22 AA593377	Human olfactory re
26	18	1.3	18	22 AAH32263	Human AFP protein
27	18	1.3	18	22 AA52209	Human prostate can
28	18	1.3	18	22 AA516217	Human polynucleoti
29	18	1.3	18	22 AA158832	Human polynucleoti
30	18	1.3	18	22 AAH14806	Human CDNA sequenc
31	18	1.3	18	22 AA41902	Nucleotide sequenc
32	18	1.3	18	22 AA41901	Nucleotide sequenc
33	18	1.3	18	22 AA160618	Human polynucleoti
34	18	1.3	18	22 AA590268	DNA encoding novel
35	18	1.3	18	22 AAC47184	Arabidopsis thalia
36	18	1.3	18	22 AB117098	Drosophila melanog
37	18	1.3	18	22 AB102222	Drosophila melanog
38	18	1.3	18	22 AA505714	Maize retinoblasto
39	18	1.3	18	22 AA505714	Maize retinoblasto
40	18	1.3	18	22 AAH18245	Human polynucleoti
41	18	1.3	18	22 AB121622	Drosophila melanog
42	18	1.3	18	22 AB121622	Human immune syste
43	18	1.3	18	22 AA542019	Genomic sequence #
44	18	1.3	18	22 AA37170	Human musculoskele
45	18	1.3	18	22 AB407858	Human ovarian and
				22 AA103669	Human reproductive

ALIGNMENTS

AAH20176	standard; DNA: 12792 BP.
AAH20176:	
09-AUG-2001 (first entry)	
Human mutated spastin nucleotide sequence SEQ ID NO:7.	
Human: mouse; spastin; ARSACS; Chromosome 13q11; Identification;	
autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;	
neurodegenerative disease; reduced sensory nerve conduction; diagnosis;	
reduced motor nerve velocity; hypermyelination of retinal nerve fibre;	
atrophy of upper cerebellar vermis; absence of Purkinje cell;	
abnormal neuronal lipid storage; genetic disorder; Characterisation; ds.	
Homo sapiens.	
Synthetic.	
Key	Location/Qualifiers
CDS	77..6604
	/*tag- a
	/product- "mutated spastin"
NO200129266-A2.	
26-APR-2001.	
20-OCT-2000; 2000WO-US29130.	
20-OCT-1999; 99US-0160588.	

Human spastin ORF
Human spastin ORF
Drosophila melanog
Novel human polynu
DNA encoding novel
C glutamicum codin
DNA encoding novel
Human polynucleoti
Human polynucleoti
Sequence encoding
Human cervical can
Human cardiovascular
C glutamicum codin
Human differential
Human polynucleoti
Human olfactory re
Human AFP protein
Human prostate can
Human polynucleoti
Human CDNA sequenc
Nucleotide sequenc
Nucleotide sequenc
Human polynucleoti
DNA encoding novel
Arabidopsis thalia
Drosophila melanog
Drosophila melanog
Maize retinoblasto
Maize retinoblasto
Human polynucleoti
Drosophila melanog
Human immune syste
Genomic sequence #
Human musculoskele
Human ovarian and
Human reproductive

XX (UYMC-) UNIV MCGILL.
PA (HOP-) HOPITAL SAINTE-JUSTINE.
XX
XX
PI Hudson TJ, Engert J, Richter A;
XX WPI: 2001-308494/32.
DR P-PSDB: AAB97821.
XX
PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -
XX
PS Claim 1: Page -: 76pp: English.

CC The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSCS)) gene sequences (I). The spastin gene has been mapped to
CC chromosome 13q11. (I) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (I) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (I). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with Spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes a mutated human spastin.
CC N.B. The present sequence is not given in the present specification but
CC is derived from the human spastin nucleotide sequence (AAB20174) as
CC stated on page 14.
XX
SQ Sequence 12792 BP: 4163 A; 2256 C; 2487 G; 3886 T; 0 other;

Query Match 100.0%; Score 1401; DB 22; Length 12792;
Best Local Similarity 100.0%; Fred. No. 0;
Matches 1401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgcagtaaaactagagcagtcgccaagcgacacaaagccttagaagaatgcatcc 60
Db 5300 gtgcagtaaaactagagcagtcgccaagcgacacaaagccttagaagaatgcatcc 5359
QY 61 aatgctgttttaacaacttggcacagaatttggcgagaagaanaattgaccacaga 120
Db 5360 aatgctgttttaacaacttggcacagaatttggcgagaagaanaattgaccacaga 5419
QY 121 attaaagatccttattatgcatatcctcttgaagaagaatgttgaagagcttccaa 180
Db 5420 attaaagatccttattatgcatatcctcttgaagaagaatgttgaagagcttccaa 5479
QY 181 aatgcgatgatcgacaagcgacagaatctgtttgttgttgccttaagacgacatca 240
Db 5480 aatgcgatgatcgacaagcgacagaatctgtttgttgttgccttaagacgacatca 5539
QY 241 gttgatagaataatttgaataaagtgtggcccatctgcaagggcgacgacttgtgtac 300
Db 5540 gttgatagaataatttgaataaagtgtggcccatctgcaagggcgacgacttgtgtac 5599
QY 301 aacaaccacgacttcaagaagaatgttgaaggaattcagaatcttggaaaaagcag 360
Db 5600 aacaaccacgacttcaagaagaatgttgaaggaattcagaatcttggaaaaagcag 5659

QY 361 aaagaggaataatccttaataaaactgagcagtatggaatgaattcattctgtatcatc 420
Db 5660 aaagaggaataatccttaataaaactgagcagtatggaatgaattcattctgtatcatc 5719
QY 421 atcacagatcgccatcttatttctgtgcaatgacatccgtgtatctttgatccatc 480
Db 5720 atcacagatcgccatcttatttctgtgcaatgacatccgtgtatctttgatccatc 5779
QY 481 gccagatagcacaaggggccacatccatagtcgccgaagcagatgtttaagatttgat 540
Db 5780 gccagatagcacaaggggccacatccatagtcgccgaagcagatgtttaagatttgat 5839
QY 541 gccagatttggacacaggttccagatgttccagatcttattcttgggaaccattttaa 600
Db 5840 gccagatttggacacaggttccagatgttccagatcttattcttgggaaccattttaa 5889
QY 601 ctgataattgcacaatgttcaagatttccctctcgttaatgacagaatgcaaaagtttc 660
Db 5900 ctgataattgcacaatgttcaagatttccctctcgttaatgacagaatgcaaaagtttc 5959
QY 661 gaaatttcgtctgttccagcatcagacagaatgttccagaatctttggacaaactgcg 720
Db 5960 gaaatttcgtctgttccagcatcagacagaatgttccagaatctttggacaaactgcg 6019
QY 721 tcagatgggcagaacttcaatgttcttcaatcacatgtgaaaaaattctatttggaa 780
Db 6020 tcagatgggcagaacttcaatgttcttcaatcacatgtgaaaaaattctatttggaa 6079
QY 781 atagataagagctatctgagcttaaatgtcgtatcagttaaagggcaaatcacagat 840
Db 6080 atagataagagctatctgagcttaaatgtcgtatcagttaaagggcaaatcacagat 6139
QY 841 gggagcagatttgaagaagaacaatttcatgcatctgttaattgatgttactcaaaag 900
Db 6140 gggagcagatttgaagaagaacaatttcatgcatctgttaattgatgttactcaaaag 6199
QY 901 agcagctcaaaagacataccagttcaacaataaccatctctgtgactgaggaactc 960
Db 6200 agcagctcaaaagacataccagttcaacaataaccatctctgtgactgaggaactc 6259
QY 961 gaaggaatcttactacgtgtgctaatttgaatagatcaggttccaagatgagga 1020
Db 6260 gaaggaatcttactacgtgtgctaatttgaatagatcaggttccaagatgagga 6319
QY 1021 gtaactaaagtgtcatatagcttcacaaagaacaaagatattcttccacaggtgt 1080
Db 6320 gtaactaaagtgtcatatagcttcacaaagaacaaagatattcttccacaggtgt 6379
QY 1081 ggaatagctgcctgcattactcaacaactaaanaaaacccataggggctctgtttttg 1140
Db 6380 ggaatagctgcctgcattactcaacaactaaanaaaacccataggggctctgtttttg 6439
QY 1141 cctcttcttcttggagactgggtgcatttcaatgtgaatgagggccacttgcagtgtatca 1200
Db 6440 cctcttcttcttggagactgggtgcatttcaatgtgaatgagggccacttgcagtgtatca 6499
QY 1201 gccagaaggaacctgtggtgtatgataatgaggtgtgttggaaatgagcttgaataac 1260
Db 6500 gccagaaggaacctgtggtgtatgataatgaggtgtgttggaaatgagcttgaataac 6559
QY 1261 agtttaatgacagatcaatagctcctgcataatgtgaatgtcaatcaatgaaataaac 1320
Db 6560 agtttaatgacagatcaatagctcctgcataatgtgaatgtcaatcaatgaaataaac 6619
QY 1321 gtaattccctgtgttctgatacaacattatcagtggttacagaacacccattatcatgt 1380
Db 6620 gtaattccctgtgttctgatacaacattatcagtggttacagaacacccattatcatgt 6679
QY 1381 taaaggaactttaaagaagt 1401
Db 6680 taaaggaactttaaagaagt 6700

RESULT 2
AAH20174
ID AAH20174 standard; DNA: 12793 BP.
XX
AC AAH20174;
XX
09-AUG-2001 (first entry)
XX
Human spastin nucleotide sequence SEQ ID NO:1.
XX
Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
KM autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KM neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KM reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
KM atrophy of upper cerebellar vermis; absence of Purkinje cell;
XX abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 77..11566
FT /tag= a
FT /product= "spastin"
XX
PN MO200129266-A2.
XX
XX 26-APR-2001.
XX
PF 20-OCT-2000; 2000MO-US29130.
XX
PR 20-OCT-1999; 99US-0160588.
XX
XX (UYMC-) UNIV MCGILL.
PA (HOP1-) HOPITAL SAINTE-JUSTINE.
XX
PI Hudson TJ, Engert J, Richter A;
XX
DR WPI: 2001-308494/32.
XX
DR P-PSDB; AAB97819.
XX
XX
PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -
XX
XX
PS Claim 1: Fig 9; 76pp; English.
XX
XX The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
CC chromosome 13q11. (I) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (I) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (I). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with Spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes human spastin as given in the present invention.
XX
XX Sequence 12793 BP; 4163 A; 2256 C; 2487 G; 3887 T; 0 other;

Query Match 92.4%; Score 1295; DB 22; Length 12793;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gtagacagtaaacactagagcagctcccaagcgacacaaagccttagaagatgacatcc 60
|||||
Db 5300 gtagcagtaaaactcagagcgcgctcccaagcgacacaaagccttagaagatgacatcc 5359
QY 61 aatgtctgttttacacacacttgccacagaaattggcgcaagaagaanaattgacacagaga 120
|||||
Db 5360 aatgtctgttttacacacacttgccacagaaattggcgcaagaagaanaattgacacagaga 5419
QY 121 attaagacatccttaatgcatatccctctgaaaagaatggtgaaagagcttctcaa 180
|||||
Db 5420 attaagacatccttaatgcatatccctctgaaaagaatggtgaaagagcttctcaa 5479
QY 181 aatgctgtagtgcgaagcgacgaagaatctgtttgtgttgatcctcagacagatcca 240
|||||
Db 5480 aatgctgtagtgcgaagcgacgaagaatctgtttgtgttgatcctcagacagatcca 5539
QY 241 gttagataaataattgataaagtgtggcccatgtgcaaggcgacactgtgtgtac 300
|||||
Db 5540 gttagataaataattgataaagtgtggcccatgtgcaaggcgacactgtgtgtac 5599
QY 301 aacacacagcatttaccagaagaatgattgaggaattcagaatcttgaaagaagcagc 360
|||||
Db 5600 aacacacagcatttaccagaagaatgattgaggaattcagaatcttgaaagaagcagc 5659
QY 361 aaagagggaaatccttataaaactgacagatggaatggaatcattctgtgtatcat 420
|||||
Db 5660 aaagagggaaatccttataaaactgacagatggaatggaatcattctgtgtatcat 5719
QY 421 atcacagactgcccactcttattctctgcaatgacatccgtgtattttgataccat 480
|||||
Db 5720 atcacagactgcccactcttattctctgcaatgacatccgtgtattttgataccat 5779
QY 481 gccagatatgcacacaggggcccacatcattagtcctcggaagcagatgtttagagatttgat 540
|||||
Db 5780 gccagatatgcacacaggggcccacatcattagtcctcggaagcagatgtttagagatttgat 5839
QY 541 gcaagatttgaagaacagtttccagatgttctgtgattcttctgtggaccattttaa 600
|||||
Db 5840 gcaagatttgaagaacagtttccagatgttctgtgattcttctgtggaccattttaa 5899
QY 601 ctggataattgacaacatgttcagatttcctctctgttaatgacgaatgtgcaaaatttcg 660
|||||
Db 5900 ctggataattgacaacatgttcagatttcctctctgttaatgacgaatgtgcaaaatttcg 5959
QY 661 gaaatttcgtctgtccagcatcagacagaatgtgtccagaatcttttggacaaactgcgc 720
|||||
Db 5960 gaaatttcgtctgtccagcatcagacagaatgtgtccagaatcttttggacaaactgcgc 6019
QY 721 tcaagatggcgagaactcttaattgttttatacacaatggaaaaaatttctattgtgaa 780
|||||
Db 6020 tcaagatggcgagaactcttaattgttttatacacaatggaaaaaatttctattgtgaa 6079
QY 781 atagataagatcacgagccttaaatgtgtgtatcagtaaaaggcaaaatccacagat 840
|||||
Db 6080 atagataagatcacgagccttaaatgtgtgtatcagtaaaaggcaaaatccacagat 6139
QY 841 ggaagcagattgaaaaggaaacaaattcagatcgttaattgatagttactataaag 900
|||||
Db 6140 ggaagcagattgaaaaggaaacaaattcagatcgttaattgatagttactataaag 6199
QY 901 aggaagcctcaagacatccagttcaaaaataaacttactatgatactgagagacct 960
|||||
Db 6200 aggaagcctcaagacatccagttcaaaaataaacttactatgatactgagagacct 6259
QY 961 gaaggaatcttactacgtgtgcttaattgtataagatcagagcttcttcaagatagagaaa 1020
|||||
Db 6260 gaaggaatcttactacgtgtgcttaattgtataagatcagagcttcttcaagatagagaaa 6319

QY	1021	gtctctaaagttcatatcagttacgaagaccgaagattatctctttccacggtgt	1080
Db	6320	gtactctaaagttgtcatatcagttacgaagaccgaagattatctctttccacggtgt	6379
QY	1081	ggagtagctgcctgcatactacacataaanaaccatagagcctctctgtttttg	1140
Db	6380	ggagtagctgcctgcattactacacataaanaaccatagagcctctctgtttttg	6439
QY	1141	cctctctctcttgagagcttggtgcattatgtgaatggtccactttgcactgtatca	1200
Db	6440	cctctctctcttgagagcttggtgcattatgtgaatggtccactttgcactgtatca	6499
QY	1201	gccagaaggaacctgtggtcgatgataatgtagtctgtgttcgaagtagtctggaataac	1260
Db	6500	gccagaaggaacctgtggtcgatgataatgtagtctgtgttcgaagtagtctggaataac	6559
QY	1261	agtttaatgacagcatataatagctcctcatatgt	1295
Db	6560	agtttaatgacagcatataatagctcctcatatgt	6594

RESULT	3
AAH20178	
ID	AAH20178 standard; DNA; 12793 BP.

DT 09-AUG-2001 (first entry)

DE Human mutated spastin nucleotide sequence SEQ ID NO:11.

KW Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
 KW autosomal recessive spastic ataxia of Charlevoix-Sauveur; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds
 XX
 OS Homo sapiens.
 OS Synthetic.

PN WO200129266-A2.

PD 26-APR-2001.

PF 20-OCT-2000; 2000WO-US29130.

PR 20-OCT-1999; 99US-0160588.

PA
(UYMC-) UNIV MCGILL.
(UOBT) UOBTAT GATEWAY BRIDGE
PA

XX
XX
ET
C
C
C
E
T
G
S
R
O
I
4
F
P
A
F
O
S

XX
XX
EFT: 2001-200404133

Now is a good time to review

PT Charlevoix-Saguenay disease by detecting two point mutations in exon 10 of the autosomal recessive spastic ataxia of

gene sequence -

Claim 1; page -; /opp; English.
XX
XX

the present invention describes human and mouse spastin, and mutated human spastin (autosomal recessive spastic ataxia of Charlevoix-Sauvé)

chromosome 13q11. (I) have neuroprotective activities and can be

gene therapy and as a spastic polypeptide agonists. (1), their fragments or their complements can be useful for assaying the presence of a nucleic

diagnosis of an early onset neurodegenerative disease in an individual.

COMPANY ACTIVE CONNECTION

CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (1) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (1). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with Spastin
CC protein or spastin gene expression or activity. The present sequence
CC represents a mutated human spastin gene from the present invention.
CC N.B. The present sequence is not given in the present specification but
CC is derived from the human spastin nucleotide sequence (AAH20174) as
CC stated on page 14.
CC
CC
XX Sequence 12793 BP; 4163 A; 2255 C 2487 G; 3888 T; 0 other;

Query Match	92.4%;	Score 1295;	DB 22;	Length 12793;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1295;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 1 gtagcagtaaaactagggagcagtcaccaagcgacacaaagccttagaaagatatgcatcc 60

Db 5300 gtagcagtaaaactagagcagtcaccaagcgacacaaagccttagaagaatatgcatcc 53:

QY 61 aatgtctgtttacaacacttggcacagaatttggcagaagaaatlgaccagcaga 12

Db 5360 aatgtctgttlaacaacacttggcacagaatttgggcagaaagaaaatltgaccagcaga 54.

121 attcaagagcatccctcaatgcatactccctctcgaagaaggaaatgctgaaagagctctctcaaa 18

DD 3420 aLlaagaycaLCCllaalgcLlCaalagaaLgcLlCaalagaaLgcLlCaal 34

[illegible][illegible]

1. The first part of the document is a list of 10 items, each consisting of a number followed by a name. The names are: 1. John Doe, 2. Jane Smith, 3. Bob Johnson, 4. Alice Brown, 5. Charlie White, 6. David Green, 7. Emily Black, 8. Frank Grey, 9. Grace Pink, 10. Henry Blue. The list is numbered 1 through 10.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1

[illegible]

Ny 421 atcaaacatccccatatctatctatccaaatnacaatccttttatcttcat

Db 5720 atcacaaactgcccatctttattctaaccaataacatcctatatttgaatcctcat 577

481 accaagatatgcaccagaqgcacatcattatgtcccgagacacatatttaagaatttgaat 540

Db 5780 qccagatatqccaccaqqccacatcattatqtcgccqacqcatgtttaagaatttqat 581

Qy 541 gcagattttagcacacagttctcagatggtctgatacttatctggaaccatttaa 601

Db 5840 gcagattttagcacagttctcagatgttcgatacttatctggaaccatttaa 585

QY 601 ctgataattgcacatgttcagatttcctctctgtaatgcagaatgycaaagtctcg 660

Db 5900 ctgataattgcacaaatgttcagatttcctccttcgtaatgcagaaatgcaaaagtctcg 5950

QY 661 gaaattcgtctgtccagcatcagacagaatggtccagaatctttgacaactgcgc 720

Db 5960 gaaattcgtctgtccagcatcagacgaatggtccagaatctttgacaaactgcgc 60:

QY 721 tcaatgagcagacacttcaatgtttcttaacatcatgtaaaatcttattgtgaa 780
DB 6020 tcaatgagcagacacttcaatgtttcttaacatcatgtaaaatcttattgtgaa 6079
QY 781 atgataagagctagtgccttaaatgtctgtatctcaagtggaaggaatcacagat 840
DB 6080 atgataagagctagtgccttaaatgtctgtatctcaagtggaaggaatcacagat 6139
QY 841 ggaagacagattgtaaaagaaacaattctcatgcatctgtatgtatgttactcaaaaag 900
DB 6140 ggaagacagattgtaaaagaaacaattctcatgcatctgtatgtatgttactcaaaaag 6199
QY 901 aggcagctcaagaacatcacacagcttcaacaataaccatactatggaatgaggaactct 960
DB 6200 aggcagctcaagaacatcacacagcttcaacaataaccatactatggaatgaggaactct 6259
QY 961 gaaagaaatcttactacgtgtgctaatgttaataagatcagagcttttcaagatggaagaa 1020
DB 6260 gaaagaaatcttactacgtgtgctaatgttaataagatcagagcttttcaagatggaagaa 6319
QY 1021 gtatctaaaggttcatatcatcagctcaagaacacacatattactctttccacgtgt 1080
DB 6320 gtatctaaaggttcatatcatcagctcaagaacacacatattactctttccacgtgt 6379
QY 1081 ggaagtaagctcctgcatctactacacatacaaaaaccataagagcttctgtttttt 1140
DB 6380 ggaagtaagctcctgcatctactacacatacaaaaaccataagagcttctgtttttt 6439
QY 1141 cctcttcttcttgagagctgtgctgcatcttcatgtgaatgagccacttgcacgtgatca 1200
DB 6440 cctcttcttcttgagagctgtgctgcatcttcatgtgaatgagccacttgcacgtgatca 6499
QY 1201 gccagaaggaactgtgtgctgtatgataatggagttgtgttcgaagtgcagtgaataac 1260
DB 6500 gccagaaggaactgtgtgctgtatgataatggagttgtgttcgaagtgcagtgaataac 6559
QY 1261 agttaatgacagcattatagctcccgcatatgt 1295
DB 6560 agttaatgacagcattatagctcccgcatatgt 6594

RESULT 4
AAH20179
ID AAH20179 standard; DNA: 12793 BP.
AC AAH20179;
XX
DT 09-AUG-2001 (first entry)
XX
DE Human mutated spastin nucleotide sequence SEQ ID NO:12.
XX
KW Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KW reduced motor nerve velocity; hypermyelination of retinal nerve fibres;
KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
XX
OS Homo sapiens.
OS Synthetic.
PN MO200129266-A2.
XX
PD 26-APR-2001.
XX
PF 20-OCT-2000; 2000MO-US29130.
XX
PR 20-OCT-1999; 99US-0160588.
XX
PA (UYMC-) UNIV MCGILL.
PA (HOP1-) HOPITAL SAINTE-JUSTINE.
XX
XX Hudson TJ, Engert J, Richter A;
PI

XX
DR WPI: 2001-308494/332.
XX
PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -
XX
PS Claim 1; Page -: 76pp; English.
XX
CC The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
CC chromosome 13q11. (I) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (I) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (I). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with Spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes a mutated human spastin.
CC N.B. The present sequence is not given in the present specification but
CC is derived from the human spastin nucleotide sequence (AAH20174) as
CC stated on page 14.
XX
SQ Sequence 12793 BP; 4163 A; 2257 C; 2487 G; 3886 T; 0 other;

Query Match 92.4%; Score 1295; DB 22; Length 12793;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtagcagtaaaactagagcagctcccaagcgacacaaagccttagaagatagcatcc 60
DB 5300 gtagcagtaaaactagagcagctcccaagcgacacaaagccttagaagatagcatcc 5359
QY 61 aatgtctgttttacaacacttgagcacaagaattggcgagaagaataattgacccgcaga 120
DB 5360 aatgtctgttttacaacacttgagcacaagaattggcgagaagaataattgacccgcaga 5419
QY 121 atgaagacatccttaatgcatatccttctgaaagaaatgttgaagaacttcttcaa 180
DB 5420 atgaagacatccttaatgcatatccttctgaaagaaatgttgaagaacttcttcaa 5479
QY 181 aatcgtatgtgcaaaagcgacagaataatcgttttgtgttgatccttagacagcatca 240
DB 5480 aatcgtatgtgcaaaagcgacagaataatcgttttgtgttgatccttagacagcatca 5539
QY 241 gtgatagaatatttgatgaagaatgggcccattggaagggccagcacttgtgtgtac 300
DB 5540 gtgatagaatatttgatgaagaatgggcccattggaagggccagcacttgtgtgtac 5599
QY 301 aacaaccagcatttacagaagaatgattgtaggaatctcgaatcttgaagaaagcagc 360
DB 5600 aacaaccagcatttacagaagaatgattgtaggaatctcgaatcttgaagaaagcagc 5659
QY 361 aaagaaggaaatccttataaaactggacagcatggaatagatcattctgtgtatcat 420
DB 5660 aaagaaggaaatccttataaaactggacagcatggaatagatcattctgtgtatcat 5719
QY 421 atcacagactgccatcttatttcttgcaatgacatccctgtgtatttgcattccat 480

```

|||||
Db 5720 atcacagactgcccacatttatttcgtgcaatgacactcgtgtattttgtaccat 5779
QY 481 gccagatgaccacagggggcacatccattagcccggaacgactgtttaagatttgat 540
Db 5780 gccagatacgacacagggggcacatccattagcccggaacgactgtttaagatttgat 5839
QY 541 gccagattttagacacagttctcagatgtctcgtatcttattcctgggaacccatttaa 600
Db 5840 gccagattttagacacagttctcagatgtctcgtatcttattcctgggaacccatttaa 5839
QY 601 ctgagataattgcaaatgtctcagatcttctcctcgttaatgacgaataatgcaaaagtctg 660
Db 5900 ctgagataattgcaaatgtctcagatcttctcctcgttaatgacgaataatgcaaaagtctg 5959
QY 661 gaatttcgtctgttcacagatcagacagatgtccagatcttctgggaacactgctgc 720
Db 5960 gaatttcgtctgttcacagatcagacagatgtccagatcttctgggaacactgctgc 6019
QY 721 tcagatggggcagaactctctaattcttctaatacacaatgcaaaatcttatttgtaa 780
Db 6020 tcagatggggcagaactctctaattcttctaatacacaatgcaaaatcttatttgtaa 6079
QY 781 atagataagagtaactgagctcctaattgtctgtatcagtaaaaggcaaatcacagat 840
Db 6080 atagataagagtaactgagctcctaattgtctgtatcagtaaaaggcaaatcacagat 6139
QY 841 gggagcaagattgaaaaagaaacaatttcacatcgtctgtaattgatagtgctactaaaag 900
Db 6140 gggagcaagattgaaaaagaaacaatttcacatcgtctgtaattgatagtgctactaaaag 6199
QY 901 aggcagctcaaaagacataccagttcaacaataactatactatgtaactgaggaactc 960
Db 6200 aggcagctcaaaagacataccagttcaacaataactatactatgtaactgaggaactc 6259
QY 961 gaaggaatcttactcgtgctgaatttgaatagatcagagcttctcaagtatggagaa 1020
Db 6260 gaaggaatcttactcgtgctgaatttgaatagatcagagcttctcaagtatggagaa 6319
QY 1021 gttctcaaaagtctcatatagctcacaagaacagatattactctttccacagtgct 1080
Db 6320 gttctcaaaagtctcatatagctcacaagaacagatattactctttccacagtgct 6379
QY 1081 ggaatgactgctcgtcatctcacaactataaaaaaaccacataggcctctctttttg 1140
Db 6380 ggaatgactgctcgtcatctcacaactataaaaaaaccacataggcctctctttttg 6439
QY 1141 cctcttcttcttgagactggtgctgcaatttcagtgaatggcacttgcactgagatca 1200
Db 6440 cctcttcttcttgagactggtgctgcaatttcagtgaatggcacttgcactgagatca 6499
QY 1201 gccacagaagaacctgctgctgtaataatgagtggtgttctggaagtgcctggaataac 1260
Db 6500 gccacagaagaacctgctgctgtaataatgagtggtgttctggaagtgcctggaataac 6559
QY 1261 agttaatgacagcatataatagctcctgcatagt 1295
Db 6560 agttaatgacagcatataatagctcctgcatagt 6594

```

RESULT 5
AAH20182
ID AAH20182 standard; DNA; 12793 BP.

XX
XX
AC AAH20182;

DT 09-AUG-2001 (first entry)

DE Human mutated spastin nucleotide sequence SEQ ID NO:15.

XX Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
KM Autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KW Neurodegenerative disease; reduced sensory nerve conduction; diagnosis;

```

KM reduced motor nerve velocity; hypermyelination of retinal nerve fibre;  
KW atrophy of upper cerebellar vermis; absence of Purkinje cell;  
KM abnormal neuronal lipid storage; genetic disorder; characterisation; ds.  
OS Homo sapiens.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT CDS 77..11566  
FT /tag= a  
FT /product= "mutated spastin"  
XX WO200129266-A2.  
XX 26-APR-2001.  
XX 20-OCT-2000; 2000MO-US29130.  
XX 20-OCT-1999; 9905-0160588.  
XX (UYMC-) UNIV MCGILL.  
PA (HOP1-) HOPITAL SAINTE-JUSTINE.  
PI Hudson TJ, Engert J, Richter A;  
PI WPI; 2001-308494/32.  
DR P-P-SDB; AAB97823.  
XX  
XX New isolated polynucleotide, encoding spastin gene, and polypeptides,  
PT useful for diagnosing autosomal recessive spastic ataxia of  
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin  
PT gene sequence -  
XX  
XX Claim 1; Page -: 76pp; English.  
PS  
XX The present invention describes human and mouse spastin, and mutated  
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay  
CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to  
CC chromosome 13q11. (I) have neuroprotective activities and can be used in  
CC gene therapy and as a spastin polypeptide agonists. (I), their fragments  
CC or their complements can be useful for assaying the presence of a nucleic  
CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the  
CC diagnosis of an early onset neurodegenerative disease in an individual.  
CC The neurodegenerative disease comprises reduced sensory nerve conduction,  
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,  
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and  
CC abnormal neuronal lipid storage. (I) can also be used to produce  
CC antisense nucleic acids, is useful as molecular weight or chromosome  
CC markers, to identify genetic disorders, as hybridisation probes or  
CC primers, as an antigen, identify and express recombinant protein for  
CC analysis, characterisation or therapeutic use, or as markers for tissues  
CC in which the corresponding protein is expressed. Diagnostic methods from  
CC the present invention can be used to identify subjects having or at risk  
CC of developing a disease or disorder associated with aberrant expression  
CC or activity of (I). The assays can be utilised to identify a subject  
CC having or at risk of developing a disorder associated with spastin  
CC protein or spastin gene expression or activity. The present sequence  
CC encodes a mutated human spastin.  
CC N.B. The present sequence is not given in the present specification but  
CC is derived from the human spastin nucleotide sequence (AAH20174) as  
CC stated on page 14.  
XX  
SO Sequence 12793 BP; 4163 A; 2257 C; 2487 G; 3886 T; 0 other;

```

Query Match 92.4%; Score 1295; DB 22; Length 12793;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtagcagtaaaactagagcagctcccaagcgacacaaagccttagaagaatagctcc 60
Db 5300 gtagcagtaaaactagagcagctcccaagcgacacaaagccttagaagaatagctcc 5359

QY 61 aatgtctgtttaaacaacttgccagaaatttggcagaaagaataatgaccagaca 120
 |||||||
 Db 5360 aatgtctgtttaaacaacttgccagaaatttggcagaaagaataatgaccagaca 5419
 QY 121 attaaagacatccttaatgcatatcctctyaaaagaaagtgtgaaagcttctcaa 180
 |||||||
 Db 5420 attaaagacatccttaatgcatatcctctyaaaagaaagtgtgaaagcttctcaa 5479
 QY 181 aatgtctgtttaaacaacttgccagaaatttggcagaaagaataatgaccagaca 240
 |||||||
 Db 5480 aatgtctgtttaaacaacttgccagaaatttggcagaaagaataatgaccagaca 5539
 QY 241 gtctgatagaatttgaataaagtgagcccatgtgcaaggccagacacttgtgtgac 300
 |||||||
 Db 5540 gtctgatagaatttgaataaagtgagcccatgtgcaaggccagacacttgtgtgac 5599
 QY 301 aacaacagacatttacaagaagatgattgtaggaatttcagaaacttgcagaaagcag 360
 |||||||
 Db 5600 aacaacagacatttacaagaagatgattgtaggaatttcagaaacttgcagaaagcag 5659
 QY 361 aaagaaggaaatcctttaaactgagacagatgataagaaatccaattcgtgtatcat 420
 |||||||
 Db 5660 aaagaaggaaatcctttaaactgagacagatgataagaaatccaattcgtgtatcat 5719
 QY 421 atccagaactgccatcttatttctgcaatgacatccgtgtattttgatacctcat 480
 |||||||
 Db 5720 atccagaactgccatcttatttctgcaatgacatccgtgtattttgatacctcat 5779
 QY 481 gccagataatgacaaaggccacataccatagtcgccgagcagatgttagaatttgat 540
 |||||||
 Db 5780 gccagataatgacaaaggccacataccatagtcgccgagcagatgttagaatttgat 5839
 QY 541 gcaaatgttgagaaacagtttcagatgatttctgatacttattctggggaaacccatttaa 600
 |||||||
 Db 5840 gcaaatgttgagaaacagtttcagatgatttctgatacttattctggggaaacccatttaa 5899
 QY 601 ctgataaattgcacaatgttccagatttccctctgtaattgacagaatggcaaaagtctcg 660
 |||||||
 Db 5900 ctgataaattgcacaatgttccagatttccctctgtaattgacagaatggcaaaagtctcg 5959
 QY 661 gaaattcgtctgttccagatcagacagaaatggtccagaaatcttcttgacaaactgagc 720
 |||||||
 Db 5960 gaaattcgtctgttccagatcagacagaaatggtccagaaatcttcttgacaaactgagc 6019
 QY 721 tcaaatgttgagaaacagtttcagatgatttctgtaattgacagaatggcaaaagtctcg 780
 |||||||
 Db 6020 tcaaatgttgagaaacagtttcagatgatttctgtaattgacagaatggcaaaagtctcg 6079
 QY 781 atagataagatgactgagacttaaatgtgtctgtatctcagtaaaaggcaaaatccagaat 840
 |||||||
 Db 6080 atagataagatgactgagacttaaatgtgtctgtatctcagtaaaaggcaaaatccagaat 6139
 QY 841 ggaagacagattgaaagaaagaaatctcaatgcatctgtaattgataagtagtactaaag 900
 |||||||
 Db 6140 ggaagacagattgaaagaaagaaatctcaatgcatctgtaattgataagtagtactaaag 6199
 QY 901 aggaagctccaagaacatacccggttcaacaataactatactatgataactgaggaactc 960
 |||||||
 Db 6200 aggaagctccaagaacatacccggttcaacaataactatactatgataactgaggaactc 6259
 QY 961 gaaagaaatcttaactagctgagcttaattgtatagatcagacttcaagatgagaagaa 1020
 |||||||
 Db 6260 gaaagaaatcttaactagctgagcttaattgtatagatcagacttcaagatgagaagaa 6319
 QY 1021 gtatctaaagatgcatatcagctcacaagaacaaagatattactcttccacgtagt 1080
 |||||||
 Db 6320 gtatctaaagatgcatatcagctcacaagaacaaagatattactcttccacgtagt 6379
 QY 1081 ggaatagctgctgctcattaccacaactataaaaaaacccttaggctctgtttttt 1140
 |||||||
 Db 6380 ggaatagctgctgctcattaccacaactataaaaaaacccttaggctctgtttttt 6439
 QY 1141 cctcttcttggagactggtcgtccatttcatgtgaatggccacttgcactgtgatcca 1200

Db 6440 cctcttcttggagactggtcgtccatttcatgtgaatggccacttgcactgtgatcca 6499
 |||||||
 QY 1201 gccagaagaaacctgtgctgcatatgataagatgtgtgttcgaagtgcagtaaac 1260
 |||||||
 Db 6500 gccagaagaaacctgtgctgcatatgataagatgtgtgttcgaagtgcagtaaac 6559
 QY 1261 agttaatgacagacttaatagctcttcataatgt 1295
 |||||||
 Db 6560 agttaatgacagacttaatagctcttcataatgt 6594
 |||||||

RESULT 6
 AAH20175
 ID AAH20175 standard; DNA; 11493 BP.
 XX
 AC AAH20175;
 XX
 DT 09-AUG-2001 (first entry)
 XX

DE Mouse spastin nucleotide sequence SEQ ID NO:3.
 XX
 KW Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
 XX
 OS Mus musculus.
 XX
 FH Key
 FT CDS 1..11493
 FT /tag= a
 FT /product= "spastin"
 XX

PN WO200129266-A2.
 XX
 PD 26-APR-2001.
 XX
 XX 20-OCT-2000; 2000WO-US29130.
 XX
 PR 20-OCT-1999; 99US-0160588.
 XX

PA (UYMC-) UNIV MCGILL.
 PA (HOPI-) HOPITAL SAINTE-JUSTINE.
 XX
 PI Hudson TJ, Engert J, Richter A;
 XX
 DR WPT: 2001-308494/32.
 DR P-PSDB: AAB97820.
 XX

PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -
 XX
 XX
 PS Claim 1; Fig 8; 76pp; English.
 XX

CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSACS)) gene sequences (1). The spastin gene has been mapped to
 CC chromosome 13q11. (1) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (1), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (1) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (1) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or

CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (I). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes mouse spastin as given in the present invention.

CC Sequence 11493 BP: 3599 A; 2281 C; 2387 G; 3226 T; 0 other:

QY Query Match 3.8%; Score 53; DB 22; Length 11493;

Best Local Similarity 100.0%; Pred. No. 8.5e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 298 tacacaacaccgaccattacagaagatgatgttagaggaattcagaattcttg 350
5521 tacacaacaccgaccattacagaagatgatgttagaggaattcagaattcttg 5573

RESULT 7

AAH20151/c
ID AAH20151 standard; DNA: 24 BP.

AC AAH20151;

DT 09-AUG-2001 (first entry)

DE Human spastin ORF PCR primer SEQ ID NO:44.

XX Human: mouse; spastin; ARSACS; chromosome 13q11; identification;
KM autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KM neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KM reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
KM atrophy of upper cerebellar vermis; absence of Purkinje cell;
KM abnormal neuronal lipid storage; genetic disorder; characterisation;
KM PCR primer; ss.

XX Homo sapiens.
OS Synthetic.

PN WO200129266-A2.

PD 26-APR-2001.

PF 20-OCT-2000; 2000WO-US29130.

PR 20-OCT-1999; 99US-0160588.

PA (UYMC-) UNIV MCGILL.
PA (HOP-) HOPITAL SAINTE-JUSTINE.

PI Hudson TJ, Engert J, Richter A;

DR WPI: 2001-308494/32.

XX New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -

PS Claim 23; Fig 7; 76pp: English.

CC The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
CC chromosome 13q11. (I) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.

CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (I) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (I). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with spastin
CC protein or spastin gene expression or activity. AAH20122 to AAH20173
CC represent specifically claimed primers which can be used in diagnostic
CC methods from the present invention.

SQ Sequence 24 BP: 8 A; 6 C; 4 G; 6 T; 0 other:

QY Query Match 1.7%; Score 24; DB 22; Length 24;

Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 793 actgagctctaaatgctgctgat 816
24 ACTGAGCTCTAAATGCTGCTGAT 1

RESULT 8

AAH20149/c
ID AAH20149 standard; DNA: 21 BP.

AC AAH20149;

DT 09-AUG-2001 (first entry)

DE Human spastin ORF PCR primer SEQ ID NO:42.

XX Human: mouse; spastin; ARSACS; chromosome 13q11; identification;
KM autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KM neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KM reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
KM atrophy of upper cerebellar vermis; absence of Purkinje cell;
KM abnormal neuronal lipid storage; genetic disorder; characterisation;
KM PCR primer; ss.

XX Homo sapiens.
OS Synthetic.

PN WO200129266-A2.

PD 26-APR-2001.

PF 20-OCT-2000; 2000WO-US29130.

PR 20-OCT-1999; 99US-0160588.

PA (UYMC-) UNIV MCGILL.
PA (HOP-) HOPITAL SAINTE-JUSTINE.

PI Hudson TJ, Engert J, Richter A;

DR WPI: 2001-308494/32.

XX New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -

PS Claim 23; Fig 7; 76pp: English.

XX The present invention describes human and mouse spastin, and mutated

CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSACS)) gene sequences (1). The spastin gene has been mapped to
CC chromosome 13q11. (1) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (1), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (1) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (1) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (1). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with Spastin
CC protein or spastin gene expression or activity. AAH20122 to AAH20173
CC represent specifically claimed primers which can be used in diagnostic
CC methods from the present invention.
XX
SQ Sequence 21 BP; 5 A; 7 C; 5 G; 4 T; 0 other;

Query Match 1.5%; Score 21; DB 22; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 gcaagggcagcacttgctg 296
Db 21 GCAAGGGCAGCAGCCTTGTGT 1
|||||

RESULT 9
AAH20150
ID AAH20150 standard; DNA; 21 BP.
XX
AC AAH20150;
XX
DT 09-AUG-2001 (first entry)
XX
DE Human spastin ORF PCR primer SEQ ID NO:43.
XX
KW Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
KW abnormal neuronal lipid storage; genetic disorder; characterisation;
KW PCR primer; ss.
XX
XX Homo sapiens.
OS Synthetic.
OS
XX WO200129266-A2.
XX
XX PD 26-APR-2001.
XX
XX PF 20-OCT-2000; 2000WO-US29130.
XX
XX PR 20-OCT-1999; 99US-0160588.
XX
XX PA (UYMC-) UNIV MCGILL.
XX (HOPIT-) HOPITAL SAINTE-JUSTINE.
XX
XX PI Hudson TJ, Engert J, Richter A;
XX
XX DR MPI; 2001-308494/32.
XX
XX PT New isolated polynucleotide, encoding spastin gene, and polypeptides,

PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -
XX
XX Claim 23; Fig 7; 76pp; English.
XX
XX The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSACS)) gene sequences (1). The spastin gene has been mapped to
CC chromosome 13q11. (1) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (1), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (1) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (1) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (1). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with Spastin
CC protein or spastin gene expression or activity. AAH20122 to AAH20173
CC represent specifically claimed primers which can be used in diagnostic
CC methods from the present invention.
XX
SQ Sequence 21 BP; 9 A; 4 C; 6 G; 2 T; 0 other;

Query Match 1.5%; Score 21; DB 22; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 gatgcaagcgacagaatc 210
Db 1 gatgcaagcgacagaatc 21
|||||

RESULT 10
AAH20152
ID AAH20152 standard; DNA; 21 BP.
XX
AC AAH20152;
XX
DT 09-AUG-2001 (first entry)
XX
DE Human spastin ORF PCR primer SEQ ID NO:45.
XX
KW Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
KW abnormal neuronal lipid storage; genetic disorder; characterisation;
KW PCR primer; ss.
XX
XX Homo sapiens.
OS Synthetic.
OS
XX WO200129266-A2.
XX
XX PD 26-APR-2001.
XX
XX PF 20-OCT-2000; 2000WO-US29130.
XX
XX PR 20-OCT-1999; 99US-0160588.
XX
XX PA (UYMC-) UNIV MCGILL.

PA (HOP1-) HOPITAL SAINTE-JUSTINE.
 XX
 PI Hudson TJ, Engert J, Richter A;
 XX
 DR WPI; 2001-308494/32.
 XX
 PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -
 XX
 PI Hudson TJ, Richter A;
 XX
 DR WPI; 2001-308494/32.
 XX
 PS Claim 23; Fig 7; 76pp; English.
 CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSCS)) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with spastin
 CC protein or spastin gene expression or activity. AAH20122 to AAH20173
 CC represent specifically claimed primers which can be used in diagnostic
 CC methods from the present invention.
 CC
 XX
 SQ Sequence 21 BP; 7 A; 5 C; 6 G; 3 T; 0 other;
 QY Query Match 1.5%; Score 21; DB 22; Length 21;
 Best Local Similarity 100.0%; Pred. No. 4.7;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 gcatcagacagatgctccag 699
 1 gcatcagacagatgctccag 21
 RESULT 11
 AAH20154
 ID AAH20154 standard; DNA; 20 BP.
 AC AAH20154;
 XX
 DT 09-AUG-2001 (first entry)
 XX
 DE Human spastin ORF PCR primer SEQ ID NO:47.
 XX
 KW Human; mouse; spastin; ARSCS; chromosome 13q11; identification;
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation;
 KW PCR primer; ss.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 WO200129266-A2.
 XX

PD 26-APR-2001.
 XX
 PF 20-OCT-2000; 2000MO-US29130.
 XX
 PR 20-OCT-1999; 99US-0160588.
 XX
 PA (UYMC-) UNIV MCGILL.
 PA (HOP1-) HOPITAL SAINTE-JUSTINE.
 XX
 PI Hudson TJ, Engert J, Richter A;
 XX
 DR WPI; 2001-308494/32.
 XX
 PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -
 XX
 PI Hudson TJ, Richter A;
 XX
 DR WPI; 2001-308494/32.
 XX
 PS Claim 23; Fig 7; 76pp; English.
 CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSCS)) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with spastin
 CC protein or spastin gene expression or activity. AAH20122 to AAH20173
 CC represent specifically claimed primers which can be used in diagnostic
 CC methods from the present invention.
 CC
 XX
 SQ Sequence 20 BP; 4 A; 5 C; 5 G; 6 T; 0 other;
 QY Query Match 1.4%; Score 20; DB 22; Length 20;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 gtgaatgcccacttgcact 1193
 1 gtgaatgcccacttgcact 20
 RESULT 12
 ABL25800/C
 ID ABL25800 standard; DNA; 8496 BP.
 AC ABL25800;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 28873.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX
 OS Drosophila melanogaster.
 OS
 XX
 WO200171042-A2.
 XX

XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001MO-US09231.
PF
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li FWD, Myers EW;
PI WPI; 2001-656860/75.
XX
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PI interactions -
XX
XX
XX Claim 1; SEQ ID NO 28873; 21pp + Sequence Listing; English.
PS
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB16351), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB57737-AB572072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 8496 BP; 2387 A; 1826 C; 1971 G; 2312 T; 0 other;
SQ

Query Match 1.4%; Score 20; DB 23; Length 8496;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 802 ctaaatgctgtatcagt 821
|||||
Db 3486 CTAATGTGCTGTATTCAGT 3467

RESULT 13
AAf66282/c
ID AAF66282 standard; cDNA; 389 BP.
XX
XX AAF66282;
AC
XX 09-APR-2001 (first entry)
DT
XX
XX Novel human polynucleotide, SEQ ID NO: 2038.
DE
XX
XX Human; cytostatic; gene therapy: colon cancer; prostate cancer;
KM breast cancer; lung cancer; cancer detection; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200102568-A2.
PN
XX
XX 11-JAN-2001.
PD
XX
XX 30-JUN-2000; 2000MO-US18374.
PF
XX
XX 02-JUL-1999; 99US-0142310.
PR 02-JUL-1999; 99US-0142311.
XX
XX (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;

PI Crkenjakov R, Drmanac S, Dickson M, Labat I, Leschkowitz D;
PI Kita D, Garcia V, Jones LM, Strache-Crain B;
XX WPI; 2001-091805/10.
DR
XX
XX Library of polynucleotides for diagnosing a cancerous state of a
PT mammalian cell and detecting cancer, particularly of the colon or
PT prostate, comprises 3351 human polynucleotide sequences -
XX
XX
XX Claim 9; Page 839; 1046pp; English.
PS
XX
XX The present sequence is one of 3351 sequences in a library of human
CC polynucleotides. The library is used to detect differentially expressed
CC genes correlated with a cancerous state of a mammalian cell and can
CC detect colon, prostate, breast and lung cancer. The library can be used
CC to produce probes for detection of mRNA and to produce additional copies
CC of the polynucleotides. The probes can be used for chromosome mapping of
CC the polynucleotide and for detection of transcription levels. Ribozymes
CC or antisense oligonucleotides can be generated. The polynucleotides and
CC their gene products are used as genetic or biochemical markers (e.g. in
CC blood or tissues) that will detect the earliest changes along the
CC carcinogenesis pathway and/or monitor the efficacy of therapies and
CC preventive interventions. The polynucleotides, polypeptides and
CC antibodies against them can be used in pharmaceutical compositions to
CC treat the cancers and proliferative disorders such as neoplasia,
CC dysplasia and hyperplasia.
XX
XX Sequence 389 BP; 107 A; 76 C; 83 G; 123 T; 0 other;
SQ

Query Match 1.4%; Score 19; DB 22; Length 389;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 402 attcaattctgtatcat 420
|||||
Db 215 ATTCAATTCTGTATTCAT 197

RESULT 14
AAS83961/c
ID AAS83961 standard; cDNA; 448 BP.
XX
XX AAS83961;
AC
XX 13-FEB-2002 (first entry)
DT
XX
XX DNA encoding novel human diagnostic protein #19765.
DE
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001MO-US08631.
PF
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Drmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
XX
XX P-PSDB; ABG19774.
DR
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess

PT biodiversity -
 XX
 PS Claim 1: SEQ ID No 19765; 103pp; English.
 CC
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AA564197-AA594564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at http://wipo.int/pub/publ/published_pcl_sequences.
 XX
 SQ Sequence 448 BP; 88 A; 129 C; 138 G; 92 T; 1 other;
 Query Match 1.4%; Score 19; DB 23; Length 448;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 550 agcacacagctccagatg 568
 |||||||
 DB 63 AGCACACAGCTTCACATG 45
 RESULT 15
 AAH66530/C
 ID AAH66530 standard; DNA; 963 BP.
 XX
 AC AAH66530;
 XX
 DT 26-SEP-2001 (First entry)
 XX
 DE C glutamicum coding sequence fragment SEQ ID NO: 1565.
 XX
 KW Corynebacterium bacterium; amino acid synthesis; vitamin; saccharide;
 XX organic acid synthesis; ds.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN EP1108790-A2.
 XX
 PD 20-JUN-2001.
 XX
 PF 18-DEC-2000; 2000EP-0127688.
 XX
 PR 16-DEC-1999; 93JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 PA (KYOWA) KYOWA HAKKO KOGYO KK.
 XX
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX
 DR WPI; 2001-376931/40.
 DR P-PSDB; AAG91311.
 XX
 PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing

PT expression profile or pattern of a gene and identifying homologous gene
 PT
 XX
 PS Claim 8; SEQ ID NO: 1565; 246pp + Sequence Listing; English.
 CC
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Corynebacterium bacterium, and identifying a homologue of a gene derived
 CC from coryneform bacterium. Corynebacterium bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 XX
 SQ Sequence 963 BP; 220 A; 261 C; 272 G; 210 T; 0 other;
 Query Match 1.4%; Score 19; DB 22; Length 963;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1271 cagcataatagctccctgc 1289
 |||||||
 DB 343 CAGCATATATAGCTCCTGC 325
 Search completed: May 22, 2002, 08:36:50
 Job time: 6630 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 08:29:03 ; Search time 91.58 Seconds
(without alignments)
3757.726 Million cell updates/sec

Title: US-09-693-205-7_COPY_5300_6700

Perfect score: 1401
Sequence: 1 gtacgacgaaactagagc.....aaagacacttaagaagt 1401

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued_Patents_NA: *
1: /cgn2_6/prodata/1/ina/5R_COMB.seq: *
2: /cgn2_6/prodata/1/ina/5R_COMB.seq: *
3: /cgn2_6/prodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/prodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/prodata/1/ina/PCrUS_COMB.seq: *
6: /cgn2_6/prodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	1.3	1806	2	US-08-819-013-2
2	17	1.2	2157	1	US-08-451-715A-3
3	17	1.2	2837	4	US-09-156-316-11
4	17	1.2	176373	3	US-09-128-155-17
5	16	1.1	20	4	US-09-467-642-57
6	16	1.1	269	3	US-08-258-287B-12
7	16	1.1	270	3	US-08-368-704C-12
8	16	1.1	345	4	US-08-591-789A-224
9	16	1.1	345	4	US-09-062-451-224
10	16	1.1	468	5	PCT-US91-08177-8
11	16	1.1	777	2	US-08-975-316-37
12	16	1.1	1299	1	US-07-688-352C-17
13	16	1.1	1299	2	US-08-474-379C-17
14	16	1.1	1299	3	US-09-146-249A-17
15	16	1.1	1299	3	US-08-206-188B-17
16	16	1.1	1299	5	PCT-US91-02714-11
17	16	1.1	1299	5	PCT-US91-02714-17
18	16	1.1	1341	4	US-09-018-635-30
19	16	1.1	1368	2	US-09-018-628-15
20	16	1.1	1368	3	US-09-273-378-15
21	16	1.1	1415	3	US-08-413-118-127
22	16	1.1	1415	3	US-08-473-446-127
23	16	1.1	1486	3	US-08-956-182-24
24	16	1.1	1511	3	US-08-956-182-41
25	16	1.1	1732	4	US-09-449-335-1
26	16	1.1	1732	4	US-09-449-335-5
27	16	1.1	1951	1	US-08-487-890A-112

28	16	1.1	1951	2	US-08-478-435-112	Sequence 112, App
29	16	1.1	1951	2	US-08-337-483-112	Sequence 112, App
30	16	1.1	1951	2	US-08-478-373-112	Sequence 112, App
31	16	1.1	1951	3	US-08-474-671-112	Sequence 112, App
32	16	1.1	1951	3	US-08-483-577A-112	Sequence 112, App
33	16	1.1	1951	4	US-08-897-438-112	Sequence 112, App
34	16	1.1	1951	4	US-08-637-654-112	Sequence 112, App
35	16	1.1	2057	3	US-09-008-103-1	Sequence 1, App1
36	16	1.1	2089	4	US-09-155-770-6	Sequence 6, App1
37	16	1.1	2119	4	US-09-018-635-28	Sequence 28, App1
38	16	1.1	2375	1	US-08-369-796-9	Sequence 9, App1
39	16	1.1	2375	2	US-08-852-091-9	Sequence 9, App1
40	16	1.1	2375	2	US-08-820-754-9	Sequence 9, App1
41	16	1.1	2375	3	US-08-956-652-9	Sequence 9, App1
42	16	1.1	2375	3	US-08-956-869-9	Sequence 9, App1
43	16	1.1	2375	3	US-08-948-547-9	Sequence 9, App1
44	16	1.1	2375	5	PCT-US95-17025-9	Sequence 9, App1
45	16	1.1	2525	4	US-09-309-487-24	Sequence 24, App1

ALIGNMENTS

```
RESULT 1
US-08-819-013-2
; Sequence 2, Application US/08819013
; Patent No. 5994522
;
; GENERAL INFORMATION:
; APPLICANT: Chan, Andrew C.
; TITLE OF INVENTION: BLNK PROTEINS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Honbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/819,013
; FILING DATE: 17-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/788,322
; FILING DATE: 24-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-64383-1/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1806 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
;
; US-08-819-013-2
;
; Query Match 1.38; Score 18; DB 2; Length 1806;
; Best Local Similarity 100.0%; Pred. No. 11;
; Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Db 1378 TTGGCGAAGAAAGAAAT 1395

RESULT 2

US-08-451-715A-3
; Sequence 3, Application US/08451715A
; Patent No. 5801013
; GENERAL INFORMATION:
; APPLICANT: Tao, Jianshi
; APPLICANT: Qiu, Yan
; APPLICANT: Houman, Fariba
; APPLICANT: Shen, Xiaoyu
; APPLICANT: Schimmel, Paul R.
; TITLE OF INVENTION: Helicobacter Aminoacyl-tRNA Synthetase
; TITLE OF INVENTION: Proteins, Nucleic Acids and Strains Comprising Same
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millitia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/451,715A
; FILING DATE: 26-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: CPT94-25
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2157 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 102..2045
; US-08-451-715A-3

Query Match 1.2%; Score 17; DB 1; Length 2157;

Best Local Similarity 100.0%; Pred. No. 36;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1385 ggacaccttaagaagt 1401

Db 185 GGACACTTTAAGAACT 201

RESULT 3

US-09-156-316-11
; Sequence 11, Application US/09156316
; Patent No. 6183961
; GENERAL INFORMATION:
; APPLICANT: Bernstein, Harold S.
; APPLICANT: Coughlin, Shaun R.
; TITLE OF INVENTION: Methods and Compositions for Regulating Cell Cycle
; FILE REFERENCE: UCSF-020/01US
; CURRENT APPLICATION NUMBER: US/09/156,316
; CURRENT FILING DATE: 1998-09-18

; EARLIER APPLICATION NUMBER: 60/060,688
; EARLIER FILING DATE: 1997-09-22
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 11
; LENGTH: 2837
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-156-316-11

Query Match 1.2%; Score 17; DB 4; Length 2837;

Best Local Similarity 100.0%; Pred. No. 36;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 684 agacagaatggtccaga 700

Db 280 agacagaatggtccaga 296

RESULT 4

US-09-128-155-17/C
; Sequence 17, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 176373
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(176373)
; OTHER INFORMATION: n = A,T,C or G
; US-09-128-155-17

Query Match 1.2%; Score 17; DB 3; Length 176373;

Best Local Similarity 100.0%; Pred. No. 35;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 751 aatcacatggaanaat 767

Db 102233 AATCACATGGAAAAAT 102217

RESULT 5

US-09-467-642-57/C
; Sequence 57, Application US/09467642
; Patent No. 6300132
; GENERAL INFORMATION:
; APPLICANT: Brett P. Monia
; APPLICANT: lex M. Cowart
; TITLE OF INVENTION: ANTISENSE MODULATION OF TLOMERIC REPEAT BINDING FACTOR 2 EXP
; CURRENT APPLICATION NUMBER: US/09/467,642
; CURRENT FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 57
; LENGTH: 20
; TYPE: DNA
; ORGANISM: Artificial Sequence

FEATURE:
OTHER INFORMATION: Antisense Oligonucleotide
US-09-467-642-57

Query Match 1.1%; Score 16; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1156 actgggctgccattc 1171
|||||

Db 20 ACTGGCTGCCATTTC 5

RESULT 6
US-08-258-287B-12
Sequence 12, Application US/08258287B
Patent No. 6083735
GENERAL INFORMATION:
APPLICANT: Yuan, Junying
TITLE OF INVENTION: Programmed Cell Death Genes and Proteins
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/258,287B
FILING DATE: 10-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/080,850
FILING DATE: 24-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Bugalsky, Lawrence B.
REGISTRATION NUMBER: 35,086
REFERENCE/DOCKET NUMBER: 0609.3920001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
TELEX: 248636 SSK
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 269 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
US-08-258-287B-12

Query Match 1.1%; Score 16; DB 3; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1366 cccctatcatgttgt 1381
|||||

Db 168 CCCCTATTCATGTGT 183

RESULT 7
US-08-368-704C-12
Sequence 12, Application US/08368704C
Patent No. 6087160
GENERAL INFORMATION:

APPLICANT: Yuan, Junying
APPLICANT: Miura, Masayuki
TITLE OF INVENTION: Programmed Cell Death Genes and Proteins
NUMBER OF SEQUENCES: 95
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/368,704C
FILING DATE: 4-JAN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/258,287
FILING DATE: 10-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
FILING DATE: 24-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Bugalsky, Lawrence B.
REGISTRATION NUMBER: 35,086
REFERENCE/DOCKET NUMBER: 0609.3920002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
TELEX: 248636 SSK
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 270 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
US-08-368-704C-12

Query Match 1.1%; Score 16; DB 3; Length 270;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1366 cccctatcatgttgt 1381
|||||

Db 168 CCCCTATTCATGTGT 183

RESULT 8
US-08-991-789A-224/C
Sequence 224, Application US/08991789A
Patent No. 6225054
GENERAL INFORMATION:
APPLICANT: Fridakis, Tony N.
Smith, John M.
Reed, Steven G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 292
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed IP Law Group
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/991,789A
APPLICATION NUMBER: US/08/991,789A
FILING DATE: 11-Dec-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Potler, Jane E. R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 210121.419C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 224:
SEQUENCE CHARACTERISTICS:
LENGTH: 345 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 224:
US-08-991-789A-224

Query Match 1.1%; Score 16; DB 4; Length 345;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 760 gaaaaattctattt 775
|||
DB 183 GAAAAATTCTATT 168

RESULT 9
US-09-062-451-224/C
Sequence 224, Application US/09062451
Patent No. 6344550
GENERAL INFORMATION:
APPLICANT: Frudakis, Tony N.
APPLICANT: Smith, John M.
APPLICANT: Reed, Steven G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 297
CORRESPONDENCE ADDRESS:
ADDRESS: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/062,451
FILING DATE: 04-APR-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.419C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 224:
SEQUENCE CHARACTERISTICS:
LENGTH: 345 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
A

US-09-062-451-224

Query Match 1.1%; Score 16; DB 4; Length 345;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 760 gaaaaattctattt 775
|||
DB 183 GAAAAATTCTATT 168

RESULT 10
PCT-US91-08177-8/C
Sequence 8, Application PC/TUS9108177
GENERAL INFORMATION:
APPLICANT: Samal, Sida K
TITLE OF INVENTION: Bovine Respiratory Syncytial Virus Genes
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESS: Venable, Baetjer, Howard & Civiletti
STREET: 1201 New York Avenue N.W., suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/08177
FILING DATE: 19911104
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/608,937
FILING DATE: 05-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Higbet, David W
REGISTRATION NUMBER: 30,265
REFERENCE/DOCKET NUMBER: 20509-96711
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4854
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 468 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bovine respiratory syncytial virus
STRAIN: A 51908
FEATURE:
NAME/KEY: CDS
LOCATION: 84..302
OTHER INFORMATION: /label= SH gene
PCT-US91-08177-8

Query Match 1.1%; Score 16; DB 5; Length 468;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 535 ttgatgcagatttta 550
|||
DB 25 TTGATGCAGATTTTA 10

RESULT 11
US-08-975-316-37
; Sequence 37, Application US/08975316
; Patent No. 5952486
; GENERAL INFORMATION:
; APPLICANT: BLOKSBERG, Leonard N., HAVUKKALA, Ilkka
; APPLICANT: and GRIERSON, Alastair W.
; TITLE OF INVENTION: MATERIALS AND METHODS FOR
; TITLE OF INVENTION: THE MODIFICATION OF PLANT LIGNIN CONTENT
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/975,316
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/713,000
; FILING DATE: September 11, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: SLEATH, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000/1003C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 777 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-975-316-37

Query Match 1.1%; Score 16; DB 2; Length 777;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 751 aatacatgtgaaaaa 766
|||
Db 755 AATCATGATGAAAAA 770

RESULT 12
US-07-688-352C-17
; Sequence 17, Application US/07688352C
; Patent No. 5527896
; GENERAL INFORMATION:
; APPLICANT: Wigler, Michael H.
; APPLICANT: Colicelli, John J.
; TITLE OF INVENTION: Cloning by Complementation and Related
; TITLE OF INVENTION: Processes
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: Two First National Plaza, 20 South Clark
; STREET: Street
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA

ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/688,352C
; FILING DATE: 19910419
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/511,715
; FILING DATE: 20-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Borun, Michael F.
; REGISTRATION NUMBER: 25447
; REFERENCE/DOCKET NUMBER: 27805/30197
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1299 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1299
; US-07-688-352C-17

Query Match 1.1%; Score 16; DB 1; Length 1299;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 960 tgaaggaatcttact 975
|||
Db 686 TGAAGGAATCTTACT 701

RESULT 13
US-08-474-379C-17
; Sequence 17, Application US/08474379C
; Patent No. 5977305
; GENERAL INFORMATION:
; APPLICANT: Wigler, Michael H.
; APPLICANT: Colicelli, John J.
; TITLE OF INVENTION: CLONING BY COMPLEMENTATION AND RELATED
; TITLE OF INVENTION: PROCESSES
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive/6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,379C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/511,715
; FILING DATE: 20-APR-1990
; PRIOR APPLICATION DATA:

```

      APPLICATION NUMBER: US 08/206,188
      FILING DATE: 01-MAR-1994
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/688,352
      FILING DATE: 19-APR-1991
      ATTORNEY/AGENT INFORMATION:
      NAME: Clough, David W.
      REGISTRATION NUMBER: 36,107
      REFERENCE/DOCKET NUMBER: 27866/32771
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (312) 474-6300
      TELEFAX: (312) 474-0448
      INFORMATION FOR SEQ ID NO: 17:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 1299 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: CDNA
      FEATURE:
      NAME/KEY: CDS
      LOCATION: 1..1299
      US-08-474-379C-17

Query Match
Best Local Similarity 100.0%; Score 16; DB 2; Length 1299;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 960 tgaaggaatcttact 975
Db 686 TGAAGGAATCTTACT 701
|||||

RESULT 14
US-09-146-249A-17
Sequence 17, Application US/09146249A
Patent No. 6069240
GENERAL INFORMATION:
APPLICANT: Migler, Michael H.
APPLICANT: Colicelli, John J.
TITLE OF INVENTION: Cloning by Complementation and Related
TITLE OF INVENTION: Processes
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/146,249A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1299 base pairs

```

```

      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: CDNA
      FEATURE:
      NAME/KEY: CDS
      LOCATION: 1..1299
      US-09-146-249A-17

Query Match
Best Local Similarity 100.0%; Score 16; DB 3; Length 1299;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 960 tgaaggaatcttact 975
Db 686 TGAAGGAATCTTACT 701
|||||

RESULT 15
US-08-206-188B-17
Sequence 17, Application US/08206188B
Patent No. 6100025
GENERAL INFORMATION:
APPLICANT: Migler, Michael H.
APPLICANT: Colicelli, John J.
TITLE OF INVENTION: Cloning by Complementation and Related
TITLE OF INVENTION: Processes
NUMBER OF SEQUENCES: 84
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/206,188B
FILING DATE: 01-MAR-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36107
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1299 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1299
US-08-206-188B-17

Query Match
Best Local Similarity 100.0%; Score 16; DB 3; Length 1299;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 960 tgaaggaatcttact 975

```

Wed May 22, 09:23:35 2002

us-09-693-205-7_copy_5300_6700.oli.rni

Page 7

|||||
Db 686 TGAAGGAATCTTACT 701

Search completed: May 22, 2002, 08:31:08
Job time: 6993 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 05:31:27 ; Search time 3619.39 Seconds
(without alignments)
2211.338 Million cell updates/sec

Title: US-09-693-205-7_COPY_12200_12792

Perfect score: 593
Sequence: 1 acatcttatttaccagct.....acctccacatttattgctt 593

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_estbm:*
3: em_estlin:*
4: em_estlmv:*
5: em_estlov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_estl2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	581	98.0	828	10	BM470780 AGENCOURT
2	565.2	95.3	660	10	BE890125
3	539.6	91.0	632	9	AM968633
4	530.6	89.5	739	9	AL567149
5	456	76.9	543	9	AI932370
6	440	74.2	536	10	BF438152
7	435.4	73.4	536	10	BE856736
8	423	71.3	500	9	AI803488
9	410	69.1	497	9	AM087745
10	374	63.1	469	9	AI217518
11	368	62.1	368	9	AA683013
12	347.6	58.6	483	10	N48291
13	341	57.5	422	9	AI078834
14	330	55.6	410	9	AA954825
15	327.4	55.2	418	10	N46342
16	323	54.5	404	9	AA809783
17	322.4	54.4	416	9	AI499896

18	310	52.3	528	9	AA417817	AA417817 zv04h08.r
19	275.2	46.4	407	9	AA481507	AA481507 aa34c04.s
20	273.6	46.1	441	10	N59442	N59442 yz30b12.s1
21	270.8	45.7	1016	10	BE896315	BE896315 601439161
22	267.8	45.2	349	9	AI561086	AI561086 tq26b04.x
23	252	42.5	351	9	AM262498	AM262498 xq85c06.x
24	240	40.5	456	9	AA417676	AA417676 zv04d08.r
25	204.4	34.5	727	10	BC619034	BC619034 602616564
26	190.2	32.1	1079	10	BM476997	BM476997 AGENCOURT
27	182	30.7	287	10	TI1045	TI1045 NIB26 Norm
28	165.2	27.9	840	10	BF693898	BF693898 602082463
29	157.2	26.5	789	10	BE889418	BE889418 601512535
30	142.4	24.0	302	9	AI865982	AI865982 wX88903.x
31	138.4	23.3	566	10	BM119091	BM119091 L0920E06-
32	138.4	23.3	711	10	BM119346	BM119346 L0926H02-
33	137.6	23.2	616	10	BM247244	BM247244 K0749B07-
34	134.4	22.7	611	9	AM555362	AM555362 L0254H06-
35	134.4	22.7	711	10	BG075163	BG075163 H3144A04-
36	132.8	22.4	543	9	AM555384	AM555384 L0255B07-
37	130.4	22.0	428	10	BE862199	BE862199 UI-M-BHO-
38	125.8	21.2	203	9	AA228047	AA228047 zY58a07.r
39	125.6	21.2	530	10	BM118416	BM118416 L0910F10-
40	119.2	20.1	589	9	AM524447	AM524447 UI-R-B00-
41	106.2	17.9	677	10	BF576822	BF576822 602132834
42	93.8	15.8	460	9	AI506773	AI506773 vM58g10.x
43	85	14.3	402	9	AV666716	AV666716 AV666716
44	83.2	14.0	464	10	BF389741	BF389741 UI-R-BS2-
45	81.8	13.8	529	9	AA124736	AA124736 m982f11.r

ALIGNMENTS

RESULT 1
BM470780 828 bp mRNA linear EST 05-FEB-2002
AGENCOURT_6463084 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5533575
LOCUS
DEFINITION
5' mRNA sequence.
ACCESSION
BM470780
VERSION
BM470780.1 GI:18519822
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
NIH-MGC http://mgc.nci.nih.gov/
1 (bases 1 to 828)
NATIONAL INSTITUTES OF HEALTH, MAMMALIAN GENE COLLECTION (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LHAM12218 row: k column: 16
High quality sequence stop: 680.
Location/Qualifiers
1. 828
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5533575"
/clone_lid="NIH_MGC_71"
/tissue_type="leptomysarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."
BASE COUNT
262 a 107 c 138 g 321 t
ORIGIN

Query Match 98.0%; Score 581; DB 10; Length 828;
 Best Local Similarity 99.7%; Pred. No. 8.9e-99;
 Matches 593; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 acatctatgtttacaggtctcctgtttgatgaagatagaacggaactcaaatggt 60
 |||||||
 DB 143 ACATCTTATAGTTTACAGGCTCTCTGTTTGATGAAGATAGCAAGCAAACTCAAAATGGT 202
 QY 61 ggcagttcttatcaacagttgttagtattgtttcttgcgaacagcttgcgaagaacatt 120
 |||||||
 DB 203 GGCAGTTCCTTATATACAGTGTGTAGTATGTTTCTGAAACTCTTGCCAGCAACATT 262
 QY 121 tattaactgttagaacaactgtcttattgttgggttatactattccacaatgttata 180
 |||||||
 DB 263 TATTAACTGTTAGACACTTGTCTTATGTTTGTGTACATATTTTCCACAAATGTTATA 322
 QY 181 attatatagttgttgaacagatgcaactcttctgttgcctaaagtgctgcaatt-- 238
 |||||||
 DB 323 ATTATATAGTGTGTGAACAGATGCAATCTTTGTTGTCTAAAGTCTGCAGTTAA 382
 QY 239 aaaaaaaaaaacactcttcttccaatagcagtgatgagtgaggttttttaacttaaaa 298
 |||||||
 DB 383 AAAAAAAAAAACACTCTTCTTCAATATGCGCATGTGAGTGTGAGCTTTTAACTTAAAA 442
 QY 299 aacacaaaattgttaaatcatctgtcttactagtagtataatataatcagcttatt 358
 |||||||
 DB 443 ACATCAAAATTTTAAATCATGTGTATCTAGTAGTTTAAATTAATATGCGCTTATAT 502
 QY 359 tccccatgaatgcatgaactgacatttaattcaatgttctgcgcagtcttcttact 418
 |||||||
 DB 503 TCCCCATGAATGATGACGAACGTGACATTTTAAATCATGTGTGTGCGCATCTTCTTACT 562
 QY 419 ttaacatattcttcttgcgaagatgaaggttaagataatagttatataagtgctact 478
 |||||||
 DB 563 TTACATATTTCTTTTGCAAAATGTAAAGTAAATATATATATTAATTAAGTGTACT 622
 QY 479 ggcgttaaatgagctcaaatatacttaataatgaaggtcctacagaacatgttgaac 538
 |||||||
 DB 623 GCGTGAATGATGCTAAATATATCTTTATGCAATTAAGGCTTACAGAACATGTTGAAC 682
 QY 539 ttttttactttatttgcgaataaagaaatgttgcacccctcacatttattgctt 593
 |||||||
 DB 683 TTTTCTTACTTTTATGGAATAGAAATGTGACCTCCACATTTTATGCTT 737
 RESULT 2
 BE890125 660 bp mRNA linear EST 20-OCT-2000
 LOCUS 601513104F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3914521 5',
 DEFINITION mRNA sequence.
 ACCESSION BE890125
 VERSION BE890125.1 GI:10348134
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 660)
 NIH-MGC http://mgi.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapds-r@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 platf: LHAM9736 row: 9 column: 02

FEATURES
 source
 High quality sequence stop: 628.
 Location/Qualifiers
 1..660
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3914521"
 /clone_1ib="NIH-MGC_71"
 /tissue_type="leiomysarcoma"
 /lab_host="PH10B (phage-resistant)"
 /note="Organ: uterus; Vector: pCMV-Sport6; Site: 1; NotI;
 Site-2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2.1 kb."

BASE COUNT 196 a 95 c 107 g 262 t
 ORIGIN

Query Match 95.3%; Score 565.2; DB 10; Length 660;
 Best Local Similarity 99.2%; Pred. No. 8.1e-96;
 Matches 589; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 1 acatctatgtttacaggtctcctgtttgatgaagatagaacggaactcaaatggt 60
 |||||||
 DB 41 ACATCTTATAGTTTACAGGCTCTCTGTTTGATGAAGATAGCAAGCAAACTCAAAATGGT 100
 QY 61 ggcagttcttatcaacagttgttagtattgtttcttgcgaacagcttgcgaagaacatt 120
 |||||||
 DB 101 GGCAGTTCCTTATATACAGTGTGTAGTATGTTTCTGAAACTCTTGCCAGCAACATT 160
 QY 121 tattaactgttagaacaactgtcttattgttgggttatactattccacaatgttata 180
 |||||||
 DB 161 TATTAACTGTTAGAACACTTGTCTTAAATGTTGTGTATGTTTCCACAAATGTTATA 220
 QY 181 attatatagttgttgaacagatgcaactcttctgttgcctaaagtgctgcaatt-a 239
 |||||||
 DB 221 ATTATATAGTGTGTGAACAGATGCAATCTTTGTTGTCTAAAGTGTCTCACTTAA 280
 QY 240 aaaaaaaaaaacactcttcttccaatagcagtgatgaggttttttaacttaaaa 299
 |||||||
 DB 281 AAAAAAAAAAACACTTCTTCTTCAATATGCGCATGTGAGTGTGTTTAACTTAAAA 340
 QY 300 catcaaaaattgttaaatatattgttattctctagtagttatataatcagcttatt 359
 |||||||
 DB 341 CATCAAAAATTTGTTAAATCATGTTATGTTATTAATTAATTAATTAATTAATTT 400
 QY 360 ccccatgaatgacgaactgacatttaattcaatgttctgcgcagtcttcttact 419
 |||||||
 DB 401 CCCCATGAATGATCAGAACGACATTTAATCATGTTGTCTGCCATGCTTCTTACTT 460
 QY 420 taacataattcttcttgcgaagtgtaaaagtgtaagtaattagttatataagtgctg 479
 |||||||
 DB 461 TAACATATTTCTTTGCGAATGTACAGGTATGATTAATTAATTAATTAATTAATGTCATG 520
 QY 480 gctgtaaatgagcttaataatattatgtgaattaaaggtctcagaacatgttgaact 539
 |||||||
 DB 521 GCTGTAATGATGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 580
 QY 540 ttttttactttatttgcgaataaagaaatgttgcacccctcacatttattgctt 593
 |||||||
 DB 581 TTTTCTTACTTTTATGCGAAT-AGGAATGTTTGCACCTCCACATTTATGCTT 633

RESULT 3
 AM968633 632 bp mRNA linear EST 01-JUN-2000
 LOCUS AM968633
 DEFINITION EST380709 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.
 ACCESSION AM968633
 VERSION AM968633.1 GI:8158474
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 632)

AUTHORS	Hedde, P., Qi R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt, I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeaman, T.J. and Quackenbush, J.
TITLE	Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray
JOURNAL	Unpublished (2000)
COMMENT	Contact: John Quackenbush The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 3528 Fax: 301 838 0208 Email: johnq@tigr.org Plate: 255 Seq primer: Forward. Location/Qualifiers 1..632 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="MAGE resequences, MAGU" /note="Vector: pBluescriptsm"
BASE COUNT	243 a 104 c 85 g 200 t
ORIGIN	
Query Match:	91.0% Score 539.6; DB 9; Length 632;
Best Local Similarity	99.1%; Pred. No. 4.7e-91;
Matches 553; Conservative 0; Mismatches 4; Indels 1; Gaps 1;	
Oy	36 atagaacgsgaaactcgaatgtggtgcagttcttatccsagtgttagtaattgttc 95
Db	632 ATGCAACGGAAATCTAATAAGTGCGCAGTCTTAATCCAGTTGTGAATGTTCT 573
Oy	96 ggaacgtcttcgcaagaacaattatthaactgtitagaacctgtcttaattgtgt 155
Db	572 GGAACGTGTCGCCAACACAATTATTAAGTAGAACACGCCTTAATGTTGTGT 513
Oy	156 gtacataattcccacaatagtataattatagtggtgtgaacagagatgaacttt 215
Db	512 GGACATATTTTCCAAATGTTATTAATTATTAAGTGTGTTGAACAGGATCCAATCTTT 453
Oy	216 ttgttctaaagtgctgcagftaaaaaaaacaaccttcttccaatagcatgta 275
Db	452 TGTTGCTTAAGTGCTGCAAGTTAAAAAAAACAACCTTTCTTCAATATGCAATGTA 393
Oy	276 gtggagttttttaacttaaacatacaaaaattgttaaatcatttgttatcagta 335
Db	392 GTGGAGTTTTTTTAAGTTAAAAAATCAAAATTTGTTAAATCATGCTTATCTAGTA 333
Oy	336 gttataatatcggtctatattccccaatgaatgatcagaactgacattaatcaatgt 395
Db	332 GTTTAT-TATTATCGGCTATATTTTCCCATGATGATCGAAGCTGACATTTAATCATGT 274
Oy	396 ttgtctgcacatgcttcttactttaacatattcctttgcgaatgtfaaaagtaatga 455
Db	273 TTGTGTGCGCAAGCTTCTTACTTTAAACATATTTTTTTTGCAGAAATGTMAAGTAATGA 214
Oy	456 taattagttatataagcttactggtctfnaatagtgcctaatacttaacttaagcaattaa 515
Db	213 TAATTAGTTTATATATAGGTACTGGCTGTAAATGATGCTAAATATATCTTTATGCAATTGA 154
Oy	516 ggagcttacagaacatgttgaaccttttactttatattggaataaagaaatgtgtcac 575
Db	153 GGAGCTTACAGAACAAGTTGAACCTTTTTTACTTTATTTGGGAATATAGGAATGTTGCAC 94
Oy	576 ctccacatttatgtctt 593
Db	93 CTCACATTTTATTATGCTT 76
RESULT 4	
AL567149/c	AL567149 739 bp mRNA linear EST 16-FEB-2001
LOCUS	AL567149 LTI-FL013.FBm1 Homo sapiens cDNA clone CSDF029YL07 3
DEFINITION	AL567149 LTI-FL013.FBm1 Homo sapiens cDNA clone CSDF029YL07 3

	prime, mRNA sequence.
ACCESSION	AL567149
VERSION	AL567149.1 GI:12920230
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 739)
AUTHORS	Li,W.B., Gruber,C., Jesse,J., and Polayes,D.
TITLE	Full-length cDNA libraries and normalization
JOURNAL	Unpublished (2001)
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - France Email: seque@genoscope.cns.fr , Web : www.genoscope.cns.fr .
FEATURES	
Source	1..739
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone="CSDF029YL07"	
/clone_idb-"LRI_Fl013_Pbm1"	
/dev_stage="pooled tissue from post conception fetuses (20 week, 24 week and 26 week)"	
/lab_host="DH10B"	
/note="Organ: Fetal brain; Vector: PCWSPORT 6; 1st strand cDNA was primed with a NotI-oli(gdT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the PCWSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies , a division of Invitrogen 9800 Medical Center Drive Rockville , Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"	
BASE COUNT	283 a 128 c 101 g 212 t 15 others
ORIGIN	
Query Match	89.5%; Score 530.6; DB 9; Length 739;
Best Local Similarity	97.6%; Pred. No. 2.2e-89;
Matches 526; Conservative 11; Mismatches 2; Indels 0; Gaps 0;	
Dy	1 acatcttatgtttaaagcgttcctglttgatbaagaatagcaaggaaaaactcaaagt 60 Db 539 ACATCTTATGTTGCAGACGGTCCTCGTTTGATCAAGTRCACAGGAATAACTAAATGT 480
OY	61 ggcaagttctaattaccagtlgttaglatglltccggaaacctgcgccgaacaacatl 120 Db 479 GGCAATTCTTAATACCAGTAGTTAATGTTCCGTTCGAAAACGCTTGCCAACACAACATT 420
OY	121 tatttaagtgtagaacactcgctttaatglttgtgtgatcatattttccaacaagtata 180 Db 419 TATTTAACGTGTGAACACTTGGCTTATGTTTTGTGTACATVTTTTGCCAATGTATFA 360
OY	181 attcataagtggttgtagacagagtgcaacttttgtgctcagaagygtcgagttaa 240 Db 359 ATTTTATATGTTGTGTTGAGCAGCATGCAATCTTTTGTGTTTAAAGGTGTCAGTTAA 300
OY	241 aaaaaaaaaaaccttttcttcaatatggcagtblagtggaagttttttaactttaaaaac 300 Db 299 AAAAAAMACMACCTTYTCTTTCAAATATGCGATGTAGTGAGTGTTTTAACTTAAAAAC 240
OY	301 atcaaaaatgtttaaatalcatatgtgttatctatggttatataaatatocgctataatlc 360 Db 239 ATTCAAAAATGTMMAAATCATATGTGTATCTAGTGATGYTAAATATATGCGCTTAATATTC 180
OY	361 ccacgaatagatcacaaacgacatlaattaaagtgttgctcgccgatgcttcttacctl 420 Db 179 CCCATGAATGATCAGAAGTIVACATTTAATTCATGTTTATTCGCGCATGCMYTCTTACTTT 120
OY	421 aacataattcttttcagaatgttaaagvtuaatgaatagttatataaagtactgacy 480

```

Db      119 AACATTTCTTTGCGAAGTGTAAAGTAATGATTAAGTTATATTAAGTGTACTCG 60
Oy      481 ctgtaaatgagctaaataactattatgcaattaaaggcttaacgaacatgttgaact 539
        |||||||
Db      59  CGTGAATGATGCTAATATATCTTTATGCAATTAAGGCTTACAGAACATGTCAMACT 1

RESULT  5
LOCUS   AI932370               543 bp      mRNA      linear      EST 17-DEC-1999
DEFINITION wd27ell.x1 Soares_NFL_T.GBC.S1 Homo sapiens cDNA clone
ACCESSION AI932370
VERSION   AI932370.1      GI:5671107
KEYWORDS EST.
SOURCE   human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 543)
AUTHORS  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
          Tumor Gene Index
JOURNAL  Unpublished (1997)
COMMENT  Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          This clone is available royalty-free through LNLN; contact the
          IMAGE Consortium (info@image.llnl.gov) for further information.
          Insert length: 982 Std Error: 0.00
          Seq primer: -400P from Gibco
          High quality sequence stop: 455.
          Location/Qualifiers
            1..543
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:232938"
              /clone_1lb="Soares_NFL_T.GBC.S1"
              /lab_host="DH10B"
              /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
              a modified polylinker; Site.1: Not I; Site.2: Eco RI;
              Equal amounts of plasmid DNA from three normalized
              libraries (fetal lung NBHL19W, testis NHT, and B-cell
              NCI-CGAP_GCB1) were mixed, and ss circles were made in
              vitro. Following HAP purification, this DNA was used as
              tracer in a subtractive hybridization reaction. The driver
              was PCR-amplified cDNAs from pools of 5,000 clones made
              from the same 3 libraries. The pools consisted of
              I.M.A.G.E. clones 297480-302087, 682632-687239,
              726408-728711, and 723096-731399. Subtraction by Bento
              Soares and M. Fatima Bonaldo."
BASE COUNT      215 a      86 c      70 g      167 t      5 others
ORIGIN

Query Match      76.9%; Score 456; DB 9; Length 543;
Best Local Similarity 99.8%; Pred. No. 1.8e-75;
Matches 467; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Oy      126 acgtttaagaacactgtcttaattgtgtgtgtacataattccacaatattataatta 185
        |||||||
Db      543 ACTGTTAGAACCTTGCTTTATGTTGTGTGTCATATTTTCCACAAATGTTATTAATTTA 484
Oy      186 tatagtgtgttgaacagatgcaatcttctgtgtctaaagtgtcgcagttaaaaaa 245
        |||||||
Db      483 TATAGTGTGTTGACAGATGCAATCTTTGTTGTCTTAAGGTCGCGCATTTAAAAAA 424
Oy      246 aaacaacctttcttcaatatagtcatgtagtgagtttttaacttaaaaaacatcaa 305
        |||||||
Db      423 AAACAACCTTTCTTCAATATGCAATGTAGTGAGTTTAACTTTAAAAAACATCAA 364
Oy      306 aaattgttaaacatgtgttctatagtagttataataataggtctatataatcccat 365
        |||||||
Db      303 AAATGTTTAAATCATGTTGTTATCTAGTAGTTTATTAATATATCGGCTTAATTTCCCAT 304
        |||||||

```

```

Oy      366 gaatgatcagaaccgacatlttaattcatgttgcgcgcagtgcttctaacttaacat 425
        |||||||
Db      303 GAATGATCAGAACGACATTTAATTCATGTTGTCTGCCCATGCTTTCTTAACAT 244
Oy      426 attcctttcagagaatgtaaaaggtaatgataatgatttatataagtgtacgtcgtga 485
        |||||||
Db      243 ATTTCTTTGCGAATGTAAAGGTAATGATTAAGTTATATTAAGTGTACTGCTGTA 184
Oy      486 aatgatcctaataacttcttgcgaattaaggcttcacagacatgttgaaccttttt 545
        |||||||
Db      183 AATGATGCTAATATATCTTTATGCAATTAAGGCTTACAGAACATGTTGAAC-TTTT 125
Oy      546 accttatttgggaataaggaatgtttgcacctccacatlttatgctt 593
        |||||||
Db      124 ACTTTTATTTGGGAATFAAGGAATGTTTGACCTCCACATTTATTTGCTT 77

RESULT  6
LOCUS   BF438152               536 bp      mRNA      linear      EST 30-MAR-2001
DEFINITION wd67f12.x1 NCI-CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3703462 3',
ACCESSION BF438152
VERSION   BF438152.1      GI:11450669
KEYWORDS EST.
SOURCE   human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 536)
AUTHORS  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
          Tumor Gene Index
JOURNAL  Unpublished (1997)
COMMENT  Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
          Emmert-Buck, M.D., Ph.D.
          CDNA Library Preparation: M. Bento Soares, Ph.D.
          CDNA Library Arrayed by: Greg Lennon, Ph.D.
          DNA Sequencing by: Washington University Genome Sequencing Center
          Clone distribution: NCI-CGAP clone distribution information can be
          found through the I.M.A.G.E. Consortium/LNLN, send email to:
          info@image.llnl.gov
          Seq primer: -400P from Gibco
          High quality sequence stop: 481.
          Location/Qualifiers
            1..536
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:3703462"
              /clone_1lb="NCI-CGAP_Lu24"
              /tissue_type="carcinoid"
              /lab_host="DH10B"
              /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
              modified polylinker; Plasmid DNA from the normalized
              library NCI CGAP_Lu5 was prepared, and ss circles were
              made in vitro. Following HAP purification, this DNA was
              used as tracer in a subtractive hybridization reaction.
              The driver was PCR-amplified cDNAs from a pool of 5,000
              clones made from the same library (clonoids
              141920-1417991 and 1520904-1522439). Subtraction by Bento
              Soares and M. Fatima Bonaldo."
BASE COUNT      212 a      85 c      67 g      172 t
ORIGIN

Query Match      74.2%; Score 440; DB 10; Length 536;
Best Local Similarity 99.8%; Pred. No. 1.7e-72;
Matches 451; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Oy      143 ttatgtttgtgtlacataattccacaatgttataattatagtggtgtgaaca 202

```


|||||
Db 536 TTATGTTGGTGTACATATTTTCCAAATGTTATATTTATATAGTGTGAGCA 477
Qy 203 ggtatgaactcttctgtcttaagtgctgcagct-aaaaaaaacaccccttctt 261
Db 476 GGATGCATCTTTTGTGTCTAAAGTGCTGCAGTTAAAAAACAACCTTTCTTT 417
Qy 262 caataagcatgtatgagagtttttaactttaaaacatcaaaatgtttaaatcat 321
Db 416 CAATATGGCATGTAGTGGAGTTTATTAACTTTAAAAACATCAAAAATGTTAAATCAT 357
Qy 322 tctgtatcagtagtataataatcagctataatcccatgaatgaacagaaactga 381
Db 356 TGCTGTTCAGTAGTATTATATATGCGCTTATATTCCCATGAATGATCAGAACGTA 297
Qy 382 caattaatcatgtctgtcgcagacgtcttcaacttaacataatcttctgcagaat 441
Db 296 CATTTAATTCATGTTTGTCTCGCATGCTCTTACTTTAACATATTTCTTTGCAGAA 237
Qy 442 gtaaaagtgatagtaattgattatataagtgctacgtgcgttaaatgataataata 501
Db 236 GTAAAGGTATGATTAATGTTTATTAATGATGCTGTAATGATGCTAAATATA 177
Qy 502 cttaatgaatlaagggtctacagaacatgtgaacacttttcaactttatgtggaata 561
Db 176 CTTATGCAANTTAAGGCTTACAGACATGTTGAACCTTTTATCTTTATGTGGAATA 117
Qy 562 aggaatgtctgcacccacatttatgctt 593
Db 116 AGGAATGTTGCACCTCCACATTTTATGCTT 85

RESULT 7
BE856736/c 536 bp mRNA linear EST 29-SEP-2000
LOCUS 7f68a06.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:3299794 3', mRNA sequence.
ACCESSION BE856736
VERSION BE856736.1 GI:10370063
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 536)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL COMMENT Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
EMAIL: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -400P from Glibco
High quality sequence stop: 448.
Location/Qualifiers
1. 336
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3299794"
/clone_id="Soares_NSF_F8_9W_OT_PA_P_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pYT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was from
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and cloneids: Soares NBHSF pool 1:
309384-310919, 323208-325895 Soares NB2HP pool 1:
145032-147335, 147720-148103, 148872-149255, 15002 -

150407, 151176-152327 Soares NB2HF8-9W pool 1:
758280-760583, 772104-774407 Soares NBHPA pool 1:
304776-306311, 320136-322823, 326280-326663 Soares NBHOT
pool 1: 723720-726407, 739080-740999 Subtraction by Bento
Soares and M. Fatima Bernaldo."
BASE COUNT 209 a 85 c 68 g 174 t
ORIGIN

Query Match 73.4%; Score 435.4; DB 10; Length 536;
Best Local Similarity 99.6%; Pred. No. 1.2e-71;
Matches 447; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 146 atgttggtgtacatatttccacaagtataattatatagtggtggaacagga 205
Db 536 ATGTGTGTGTACATATTTTCCAAATGTTATATTTATATAGTGTGAGCA 477
Qy 206 tgcatactcttgctcctaaagtgctgcagct-aaaaaaaacaccccttcttcaa 264
Db 476 TGCAATCTTTTGTGTCTAAAGTCTCTGACGTTAAAAAACAACCTTTCTTTCA 417
Qy 265 tatgcatgtatgtgagtttttcaacttaaaacatcaaaatgtttaaatcatgt 324
Db 416 TATGCAATGATGAGGTTTATTAATGATGCTGTAATGATGCTAAATATGCT 357
Qy 325 gtactcagtgattataataatcgcctataatcccatgaatgaacagacgtacat 384
Db 356 GTATCTGTATGATTAATTAATTAATGCTTATTTCCCATGAATGATGATGATG 297
Qy 385 ttaactcagtgctgcgcacagctcttcaacttaacataatcttcttggaagatga 444
Db 296 TTAATTCATGTTTGTCTGCGCATGCTCTTACTTTAACATATTTCTTTGCAATGA 237
Qy 445 aaaggtaatgaatlaattatataatagtgctgcgtgtaaaatgataataatct 504
Db 236 AAAGTATGATTAATTAATTAATTAATGATGCTGCTGAATGATGCTAAATATGCT 177
Qy 505 tatgcaatlaagggtctacagaacatgtgaacacttttcaacttttcttggaatga 564
Db 176 TATGCAATTAAGGCTTACAGAACATGTTGAACCTTTTACTTTATTTGGAATMAG 117
Qy 565 aatgttgacccacacatttatgctt 593
Db 116 AATGTTTGCACCTCCACATTTTATGCTT 88

RESULT 8
A1803488/c 500 bp mRNA linear EST 13-DEC-1999
LOCUS tc17g02.x1 Soares_NHMPu_S1 Homo sapiens cDNA clone IMAGE:2064146
DEFINITION 3', mRNA sequence.
ACCESSION A1803488
VERSION A1803488.1 GI:5368882
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 500)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL COMMENT Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
EMAIL: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1089 Std Error: 0.00
Seq primer: -400P from Glibco
High quality sequence stop: 447.
Location/Qualifiers
1. 500
/organism="Homo sapiens"

FEATURES
source

/db_xref="taxon:9606"
 /clone.lib="Soares.NhHMPU.S1"
 /tissue_type="Pooled human melanocyte, fetal heart, and pregnant uterus"
 /lab_host="DH10B"
 /note="Organ: mixed (see below); Vector: pT73D-Pac (pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NBHM, pregnant uterus NBHPU, and fetal heart NBH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

BASE COUNT 196 a 80 c 62 g 162 t
 ORIGIN

Query Match 71.3%; Score 423; DB 9; Length 500;
 Best Local Similarity 100.0%; Pred. No. 2.5e-69;
 Matches 423; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 aatgtataattatagtggtgagcagatcattctgtgctaaagtg 230
 |||||||
 Db 500 AATGTTATATTATATAGTGTGGAACAGATCAATCTTTGTGCTAAAGTG 441
 QY 231 ctgcagtttaaaaaaacacttcttccaatgcatgtagtggagtttttaa 290
 |||||||
 Db 440 CTCAGTTAAAAAACAACCTTTCTTCAATATGATGATGAGGATTTTAA 381
 QY 291 cttaaaaacataaaatgtaaaatcattggtatctagtaagtttaatacgg 350
 |||||||
 Db 380 CTTTAAAAATCAAAATGTTAAATCATGTGTTATGATGTTAAATTAACGG 321
 QY 351 ctatacttcccatgaatgacagactaattcaatcattgtctcgcagatc 410
 |||||||
 Db 320 CTTATATTTCCCATGATGATGATGATGATGATGATGATGATGATGATG 261
 QY 411 tcttacccttaacatattcttctgcagaatgtaaaagtaataagttatata 470
 |||||||
 Db 260 TCTTTACTTTAAATATTTCTTTGCGAATGTAAAGTAATGATTAATTAATA 201
 QY 471 agtgaactgctgtaaatgctaaatataacttaagaaggtctaacgaacat 530
 |||||||
 Db 200 AGGTACTGCTGTAAATGATGATGATGATGATGATGATGATGATGATGAT 141
 QY 531 gtgaacacttttcttacttactggaataagaatggttgcacccacatttatg 590
 |||||||
 Db 140 GTGAACCTTTTACTTTATTTGGAATTAAGGAATGTTTGCACCTCCACATTTTATG 81
 QY 591 ctt 593
 |||
 Db 80 CTT 78

RESULT 9
 AM087745/c 497 bp mRNA linear EST 15-OCT-1999
 LOCUS x66f08.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
 DEFINITION IMAGE:2581479 3', mRNA sequence.
 ACCESSION AM087745
 VERSION AM087745.1 GI:6043550
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 497)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL Tumor Gene Index,
 COMMENT Unpublished (1997)
 CONTACT Robert Strausberg, Ph.D.
 EMAIL: cgaps-remail.nih.gov
 THIS CLONE IS AVAILABLE ROYALTY-FREE THROUGH LNL; CONTACT THE
 IMAGE CONSORTIUM (info@image.llnl.gov) FOR FURTHER INFORMATION.
 SEQ PRIMER: -400P from Glibco
 HIGH QUALITY SEQUENCE STOP: 455.
 LOCATION/QUALIFIERS
 1..497
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone.lib="Soares.NhHMPU.S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pT73D-Pac (pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBH19W, testis NHT, and B-cell NCI-CGAP GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Benito Soares and M. Fatima Bonaldo."

BASE COUNT 195 a 80 c 63 g 159 t
 ORIGIN

Query Match 69.1%; Score 410; DB 9; Length 497;
 Best Local Similarity 99.8%; Pred. No. 6.6e-67;
 Matches 421; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 172 aatgtataattatagtggtgagcagatcattctgtgctaaagtg 231
 |||||||
 Db 497 AATGTTATATTATATAGTGTGGAACAGATCAATCTTTGTGCTAAAGTG 438
 QY 232 tgcagtttaaaaaaacacttcttccaatgcatgtagtggagtttttaa 291
 |||||||
 Db 437 TCAGTTAAAAAACAACCTTTCTTCAATATGATGATGAGGATTTTAA 378
 QY 292 cttaaaaacataaaatgtaaaatcattggtatctagtaagtttaatacgg 351
 |||||||
 Db 377 TTTAAAAATCAAAATGTTAAATCATGTGTTATGATGATGATGATGATGAT 318
 QY 352 ttatattcccatgaatgacagactaattcaatcattgtctcgcagatc 411
 |||||||
 Db 317 TTATATTTCCCATGATGATGATGATGATGATGATGATGATGATGATGAT 258
 QY 412 cttaacttcaacatattcttctgcagaatgtaaaagtaataagttatataa 471
 |||||||
 Db 257 CTTTACTTTAAATATTTCTTTGCGAATGTAAAGTAATGATTAATTAATA 198
 QY 472 gtgaactgctgtaaatgctaaatataacttaagaaggtctaacgaacatg 531
 |||||||
 Db 197 GTGACTGCTGTAAATGATGATGATGATGATGATGATGATGATGATGATG 138
 QY 532 ttgaacacttttcttacttactggaataagaatggttgcacccacatttatg 591
 |||||||
 Db 137 TTGAACCTTTTACTTTATTTGGAATTAAGGAATGTTTGCACCTCCACATTTTATG 79
 QY 592 tt 593
 |||
 Db 78 TT 77

RESULT 10
 AI217518/c 469 bp mRNA linear EST 17-MAR-1999
 LOCUS AI217518
 DEFINITION qn20g08.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
 IMAGE:1845278 3', mRNA sequence.

ACCESSION AI217518
VERSION AI217518.1 GI:3797333
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 469)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1085 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 456.
Location/Qualifiers
1..469
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1845278"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled. Vector: pT773D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI-CCAP GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 184 a 71 c 63 g 151 t
ORIGIN
Query Match 63.1%; Score 374; DB 9; Length 469;
Best Local Similarity 99.7%; Pred. No. 3.3e-60;
Matches 385; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 209 aatccttctgtctaaagtgctgcagtttaaaaaa-caaccttcttcaatat 267
|||||
DB 469 AATCTTTGTTGCTTAAGGTCGCGAGTTAAAAAACCAACCTTTCTTCAATAT 410
QY 268 ggcacgtcagtgagcttcttaacttaaaaacatcaaaaatgtttaaatcattgtgt 327
|||||
DB 409 GGCATGTAGTGGAGTTTAACTTTAAAAACATCAAAAAATGTTAAATCATATGTGCTT 350
QY 328 atcgaagatgtaataatcagcttatattcccccagatgctcagaactgacattta 387
DB 349 ATCTAGAGATTATTAATTAATGCTTATATTTCCCATGAATGATCAACATGACATTTA 290
QY 388 attcaatgttctcgcacatgctcttacttaacatattcttcttcagaatgtaaa 447
|||||
DB 289 ATTCATGTTTGTCGGCAGTCCTTACTTTAAACATATTTCTTTTCAGAGATTAATA 230
QY 448 ggtataataatgattatataaagtgactgctgtaaatgataatataacttat 507
DB 229 GGTAAATATATATGTTTATTAAGTACTGCTGTAAGATGCTAAATATATATCTTAT 170
QY 508 gcaattagggctcagaagaacattgaacttttacttatttgggaataagat 567
DB 169 GCATTAAGGCTTACAGAACATGTTGAAACTTTTCTTACTTTATTTGGAAATTAAGAT 110
QY 568 gtctgcacctcacatttattgtct 593
DB 109 GTTGCACCTCCACATTTTATGCTT 84

RESULT 11
AA683013/c 368 bp mRNA linear EST 15-DEC-1997
LOCUS
DEFINITION aeg1b08.s1 Stratiogene schizo brain S11 Homo sapiens cDNA clone
IMAGE:970551 3', mRNA sequence.
ACCESSION AA683013
VERSION AA683013.1 GI:2668904
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 368)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Giesel, G., Jost, S.,
Kizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin,
J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
White, Y., Wylie, T., Waterston, R. and Wilson, R.
TITLE Unpublished (1997)
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estevenson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: polyt not found
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 358.
Location/Qualifiers
1..368
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:970551"
/clone_lib="Stratiogene schizo brain S11"
/sex="male"
/tissue_type="schizophrenic brain S-11 frontal lobe"
/dev_stage="34 years old"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: Bluescript SK-; Site_1: EcoRI. Library
constructed from S-11 frontal lobe, male, 34 years old,
50% caucasian, 50% Aleutian. Schizophrenic suicide.
Random primed into EcoRI site of ZAP II Vector. Mass
excised. Custom library. Avg insert length 1.4kb.
Material obtained by Johnston N., Torrey, E.F., Yolken R.,
and the Stanley Neuropathology Consortium - Analysis of
RNAs from the Brains of Individuals with Psychiatric
Diseases (unpublished) Stanley Neurovirology Laboratory,
Johns Hopkins School of Medicine, Baltimore MD."
BASE COUNT 140 a 61 c 52 g 115 t
ORIGIN
Query Match 62.1%; Score 368; DB 9; Length 368;
Best Local Similarity 100.0%; Pred. No. 4.6e-59;
Matches 368; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 35 gatacaacggaaactcaaatgtagtgagcttctataccagttgtagatgtttc 94
DB 368 GATACCAACGGAAACTCAAAATGTGGCAGTTCTTATTAACAGTTGATGATGTTTC 309
QY 95 tgaagctgctgcgaagaacatttataactgcttgaagaactgcttcttggtg 154
DB 308 TCGAAGCTGCTGCCAAGCAACATTATTAATCTGTTAAACACTGCTTATGTTGG 249
QY 155 tgtacatatcttcacaatgltataatlatatagtggtgtggaacagatgcaatc 214
DB 248 TGTACATATTTTCCACAAATGTTATTAATTAATGATGTTGAACAGATGCAATCTT 189
QY 215 ttgtgtctaaagtgctgcagtttaaaaaaaacacttcttctcaatatgacatgt 274

```

|||||
Db 188 TTGTTGCTAAGTCTGCAGTAAACCAACCTTTCTTCAATATGCGATGT 129
OY 275 agtggagtttttaacttaaaacatacaaaatgtttaaacattggttctagt 334
Db 128 AGTGGAGTTTAACTTTTAAAAACATCAAAATTTGTTAAATCATGTGTATGTAGT 69
OY 335 agttataatcatcgcctatattcccccatagaatcagaactgacattatcatg 394
Db 68 AGTTTAAATTAATTCGGCTTATATTTTCCCATGATGATGATGACACTGATTAATTCATG 9
OY 395 ttgtgtctc 402
Db 8 TTGTCTC 1

RESULT 12
M48291/c 483 bp mRNA linear EST 14-FEB-1996
LOCUS N48291
DEFINITION yy77c12.s1 Soares_multiple_sclerosis_2NbHSP Homo sapiens cDNA
ACCESSION N48291
VERSION N48291.1 GI:1189457
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 483)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Maira,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Treviskis,E., Waterston
,R., Williamson,A., Woldmann,P. and Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
JOURNAL
COMMENT Contact: Wilson RK
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGe Consortium (info@image.llnl.gov) for further information.
Seq primer: ml3 -40 forward
High quality sequence stop: 205.
Location/Qualifiers
1. 483
/organism="Homo sapiens"
/db_xref="GDB:3897950"
/db_xref="taxon:9606"
/clone="IMAGE:279574"
/clone_lib="Soares_multiple_sclerosis_2NbHSP"
/sex="male"
/tissue_type="multiple_sclerosis_lesions"
/dev_stage="Age 46"
/lab_host="DH10b (ampicillin resistant)"
/notes="Vector: pT7T3D (Pharmacia) with a modified
polylinker V-type phagemid; Site_1: Not I; Site_2: Eco RI
; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5'
TGTTCACCAATCTGAAGTGGAGCGCGCATTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Falima Bonaldo. RNA from 4 multiple sclerosis
lesions from one patient was kindly provided by Dr. Kevin
G. Becker (NINDS/NIH)."
BASE COUNT 177 a 80 c 71 g 147 t 8 others
ORIGIN

```

```

Query Match 58.6%; Score 347.6; DB 10; Length 483;
Best Local Similarity 93.6%; Pred. No. 2,76-55;
Matches 380; Conservative 0; Mismatches 22; Indels 4; Gaps 2;

OY 192 gtggttgaacagatgcatctttgtgtctaaagtgctgcagttaaaaaaacaa 251
Db 482 GGGNTGGACNCGATGCCATCTTTGTGNCNCAAGGTGCGCCAGTTAAAAAACCC 423
OY 252 c---tttcttcaatagc-catgtagtgaggttttttcttaactttaaacatcaaa 307
Db 422 ACCCTTTCTTCAATATAGCCATGTAGTGAGCTTTTAACTTNAACATCAAAA 363
OY 308 atgttaaaatcattgtgtatcctagtagttataatcatcgcctatattcccatga 367
Db 362 ATGTGTAATATCATGTGTATCTAGNAGTTAATAATATGCGGCTTATATTTCCCATGA 303
OY 368 atgatacgaactgacattatcatgttctgcgcacatgctctcttaacttaacata 427
Db 302 ATGATCAGAACTGACATTAATTAATTCATGTTCNCGCATCTCTTAACTTAACATAT 243
OY 428 ttctttgcagaatgtaaaaggtatgtaattgttataatagtgactgctgttaa 487
Db 242 NCTTTTGCAGAAATGTAAGGTAATGTAATGTTATATTAAGTACTGCTGTAAA 183
OY 488 tgatgctaataatacttataatgaatgaaggcttacagaacatgttgaactttttac 547
Db 182 TGATGCTAATATATCTTATATGCAATTAAGGCTTACACAAACATGTTAACTTTTAC 123
OY 548 tttaattgggaataaggaatgttgcacctccacatttattgctt 593
Db 122 TTTTATGGGAATTAAGAAATGTTTGACCTTCACACTTTTATTTGCTT 77

RESULT 13
A1078834/c 422 bp mRNA linear EST 10-AUG-1998
LOCUS A1078834
DEFINITION 0246d05.x1 Soares_NbHMPu_S1 Homo sapiens cDNA clone IMAGE:1678377
ACCESSION A1078834
VERSION A1078834.1 GI:3413141
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 422)
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-rt@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGe Consortium (info@image.llnl.gov) for further information.
Seq primer: -40ml3 fwd. EV from Amersham
High quality sequence stop: 361.
Location/Qualifiers
1. 422
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1678377"
/clone_lib="Soares_NbHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10b"
/notes="Organ: mixed (see below); Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbHMP, pregnant uterus
NbHMP, and fetal heart NbH119n) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of

```


Gencore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 06:53:41 ; Search time 3530.57 Seconds
(without alignments)
3514.853 Million cell updates/sec

Title: US-09-693-205-7_COPY_12200_12792

Perfect score: 593

Sequence: 1 acatctatgtttacagct.....acctccacatttattgctt 593

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_bhg: *
3: gb_in: *
4: gb_ov: *
5: gb_ov: *
6: gb_ph: *
7: gb_ph: *
8: gb_pl: *
9: gb_pl: *
10: gb_ro: *
11: gb_ro: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_mu: *
21: em_ov: *
22: em_ov: *
23: em_ph: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_ro: *
28: em_un: *
29: em_vl: *
30: em_hhg_hum: *
31: em_hhg_inv: *
32: em_hhg_other: *
33: em_hhg_inv: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query No. Score Match Length DB ID Description

1	593	100.0	4318	9	AB018273	AB018273 Homo sapi
2	593	100.0	12793	6	AX119931	AX119931 Sequence
3	593	100.0	12793	9	AF193556	AF193556 Homo sapi
4	593	100.0	92693	9	AL157766	AL157766 Human DNA
5	593	100.0	99819	2	AC079761	AC079761 Homo sapi
6	426.6	71.9	3289	9	AB056815	AB056815 Macaca fa
7	327.4	55.2	418	11	G36555	G36555 SHCC-53325
8	61	10.3	169794	2	AC004688	AC004688 Plasmodu
9	61	10.3	196149	2	AC004709	AC004709 Plasmodu
10	59.6	10.1	4601	3	DMU11584	U11584 Drosophila
11	59.6	10.1	19517	3	DMU37541	U37541 Drosophila
12	57.4	9.7	12029	3	AE001426	AE001426 Plasmodu
13	57.2	9.6	60604	2	AC023466	AC023466 Homo sapi
14	57	9.6	183813	2	AC012204	AC012204 Homo sapi
15	57	9.6	197225	2	AC093835	AC093835 Homo sapi
16	56.8	9.6	110000	2	PFMAL4P1_2	Continuation (3 of
17	56.6	9.5	11691	6	AX347143	AX347143 Sequence
18	56.6	9.5	349980	6	AX344564	AX344564 Sequence
19	55.4	9.3	124635	9	AP000593	AP000593 Homo sapi
20	55.2	9.3	1815	1	AF267220	AF267220 Candidatu
21	55.2	9.3	162075	9	HS127D3	AL021026 Homo sapi
22	55	9.3	1141	6	AX083744	AX083744 Sequence
23	55	9.3	13038	6	AX346176	AX346176 Sequence
24	55	9.3	145598	9	AC008132	AC008132 Homo sapi
25	54.8	9.2	15387	6	AX345086	AX345086 Sequence
26	54.6	9.2	4611	3	PFAL32006	AL1332006 Plasmodu
27	54.4	9.2	7918	8	AF325123	AF325123 Arabidops
28	54.4	9.2	13574	6	AX346218	AX346218 Sequence
29	54.4	9.2	62268	9	HS1178121	AL109852 Human DNA
30	54.4	9.2	141748	8	AC009323	AC009323 Arabidops
31	54.4	9.2	187048	31	AC024921	AC024921 Homo sapi
32	54	9.1	349980	6	AX344558	AX344558 Sequence
33	54	9.1	349980	6	AX344559	AX344559 Sequence
34	53.8	9.1	863	11	CNS06EVO	AL395628 T7 end of
35	53.8	9.1	189790	2	AC107420	AC107420 Homo sapi
36	53.4	9.0	186625	9	AC006487	AC006487 Homo sapi
37	53	8.9	141469	9	AL139811	AL139811 Human DNA
38	53	8.9	161384	2	AC083801	AC083801 Homo sapi
39	52.8	8.9	17848	6	AX277865	AX277865 Sequence
40	52.8	8.9	17848	6	AX323550	AX323550 Sequence
41	52.8	8.9	17848	6	AX348363	AX348363 Sequence
42	52.8	8.9	19087	6	AX345694	AX345694 Sequence
43	52.8	8.9	77835	2	PFMAL13P2_3	Continuation (4 of
44	52.6	8.9	150754	9	AC023491	AC023491 Homo sapi
45	52.6	8.9	175358	9	AC007981	AC007981 Homo sapi

ALIGNMENTS

RESULT 1
AB018273 4318 bp mRNA linear PRI 16-JUN-1999
LOCUS Homo sapiens mRNA for KIAA0730 protein, partial cds.
DEFINITION AB018273
ACCESSION AB018273
VERSION AB018273.1 GI:3882180
KEYWORDS
SOURCE Homo sapiens adult male brain cDNA to mRNA, clone_11b:pluascript11
SK plus clone:HK03632.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Miyajima,N.,
Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes. XI.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro
JOURNAL DNA Res. 5 (5), 277-286 (1998)
MEDLINE 99087487
REFERENCE 2 (bases 1 to 4318)
AUTHORS Ohara,O., Suyama,M., Nagase,T., Ishikawa,K. and Kikuno,R.
TITLE Direct Submission

JOURNAL Submitted (08-OCT-1998) Osamu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology, Yana 1532-3, Kisarazu, Chiba 292-0812, Japan (E-mail:cdna@info.kazusa.or.jp, Tel:+81-438-52-3913, Fax:+81-438-52-3914)

FEATURES
source Location/Qualifiers

1.4318
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hK03632"
/sex="male"
/tissue_type="brain"
/clone_1ib="pBluescriptII SK plus"
/dev_stage="adult"
1.3015
/gene="KIAA0730"
<1.3015
/gene="KIAA0730"
/codon_start=1
/product="KIAA0730 protein"
/protein_id="BA034450.1"
/db_xref="GI:3882181"
/translation="VTEMTSWVEFLRNIGLKLTLSQOQLDPKAEISVANENNSKE
TQNTVDILHHI FGRMDLSGNLEKLSLPFLCPKPAPELIFHHOYDEVNGL
PLIKNGAGVNPKEKQDVLOLMTSCPLPKAPLSIKEGSDLGQOEOLEVLN
MLNVLDPLDKVINNCRNICITITDEEMVTRAKVLSRISLEAEKREFEFLRG
VAFVVEDMKLLEPEVYINLEYESDFPYLYKLDELGTFOHLEKHGTEDIISTK
OYVEVLSRIEKGSEKQLODPNEMRYKRVSLFRSLQSDSVYKRDLDLVNVDLXYL
PSODGRUYSSILVDPDAPHYKRSIOGNIGVOMVLDSCYLGKDKGFEKTLMEFQ
KLRPLLSILEQDEETPKVQGFALDSLRGLDLSLEDFITGLRIMKHENDN
AFLANEKAIRLCKALREGLKVCSEFQLQTLTRKGFNP1PHSRSTFPLKFGNAV
ILLYIOHSDSKINFLALAMLKSAIDMLISPTXYLIAMLCNDYRIRGEKIDSLG
KYDSEPSKLELPMPTPIPAEIHITLMDPMNVFEGEVGYLVDAEGDIDYSGOP
TYTVYIIVOEVEDADNNSPFGKIYOIDIGYSEKVIASLIDIKFSPREESQSDRS
ASTPSPTEFLTPGLRSTPLPFSGRESKTSKHSKPKLKVNSIPELIKETSVYE
QAMKLPESERKKIIRRLYLKMPDKNPENHDIANEFKHLQNTINLEQAFIDQAD
RASRTEFSTSASRFQSDKYSFQRFYTSNMOEATSHKSEKQONKEKCPSSAGQYSGR
FEVPTPFKSVNGVEARMLROARNFSAANDLHNANEMWCFKCYLSLKLALLAD
YAVRGSDKDVPTALAKRIEYSQLELITNDVHTLEAVGYDSLSTRYPDLLPFQI
PNDRFSEVAMVMECTACIIITIKLENMQKV"

BASE COUNT 1395 a 734 c 847 g 1342 t
ORIGIN

Query Match 100.0%; Score 593; DB 9; Length 4318;
Best Local Similarity 100.0%; Pred. No. 4,7e-92;
Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acatcttaagttaacaggtctcctgttgatgaagatgaacaggaacccaatggt 60
|||||
DB 3650 ACATCTTAAGTTTAAACAGGCTTCTGTTGATGAAGATGCAAGAAACTCAAAATGCT 3709
QY 61 ggcagttcttataccaggtgttagtattgttcggaacgtctgccaagaacatt 120
|||||
DB 3710 GGCAGTCTTATTAACAGGTTGTAGTATGTTTCTGGAACCTGCTGCCAAGCAACATT 3769
QY 121 tataactgttagaacactgtcttattgttggtagacatttccacaactgttata 180
|||||
DB 3770 TATTAACTGTTAACACACTTGTATTGTTGTGTGACATATTTTCCAAATGTTATA 3829
QY 181 attatataagtgtagtgaacagatgcaatcttctgtgtctaaagtgctcagttaa 240
|||||
DB 3830 ATTATATATAGTGTGTGAACAGATGCAATCTTTGTGTCTAAAGTGCTGCAGTTAA 3889
QY 241 aaaaaaaacaactttcttccaatagatgcatgtagtgaggtttttaaacttaaaac 300
|||||
DB 3890 AAAAAAAACACCTTTCTTCAATATGCGATGTAGGAGGTTTCTTAACTTTAAATAC 3949
QY 301 atcaaaaatgttaaaaatcaatctgttatctagtagttaattatcgcttatattc 360
|||||
DB 3950 ATCAAAAATGTGTTAAATCATCTGTGTATCTAGTAGTTTAAATATCGCCTTATATTC 4009
QY 361 cccatgaatgaatgaacagactgaattaatcatggttgcctgcacatgcttcttactt 420
|||||

DB 4010 CCCATGATGATGACAGACTGACATTTAATTCATGTTGTCTGCCCATCTTCTTACTTT 4069
QY 421 aacatattcttttgcgaatgtaaaagtgtaagataatagttatataagtgctcg 480
|||||
DB 4070 AACATATTTCTTTTGGCAGATGTAAGGTAATGATATTAATTAATAGTGTACTGG 4129
QY 481 ctgtaaatgtcttaaatataacttataatgcaattaaaggtcttaagaacatgctgaaact 540
|||||
DB 4130 CTGTAATATGATGCTAATATATACCTTTATGCAATTAAGGCTTACAGAACTGTGAACCTT 4189
QY 541 ttttacttatttgggaataaggaatglttgcaacctccacatttatgctt 593
|||||
DB 4190 TTTTACTTTATTTGGGAATAGGAATGTTTGGACCTCCACATTTATTTGCTT 4242

RESULT 2
AX119931
LOCUS AX119931 12793 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 1 from Patent WO0129266.
ACCESSION AX119931
VERSION AX119931.1 GI:14036678
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE
1 (bases 1 to 12793)
Hudson,T.J., Engert,U. and Richter,A.
AUTHORS Identification of arcs mutations and methods of use therefor
TITLE Patent: WO 0129266-A 1 26-APR-2001;
JOURNAL MCGILL UNIVERSITY (CA); Hopital Sainte-Justine (CA)
LOCATION/Qualifiers

FEATURES
source 1.12793
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 4163 a 2256 c 2487 g 3887 t
ORIGIN

Query Match 100.0%; Score 593; DB 6; Length 12793;
Best Local Similarity 100.0%; Pred. No. 3.6e-92;
Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acatcttaagttaacaggtctcctgttgatgaagatgaacaggaacccaatggt 60
|||||
DB 12201 ACATCTTAAGTTTAAACAGGCTTCTGTTGATGAAGATGCAAGAAACTCAAAATGCT 12260
QY 61 ggcagttcttataccaggtgttagtattgttcggaacgtctgccaagaacatt 120
|||||
DB 12261 GGCAGTCTTATTAACAGGTTGTAGTATGTTTCTGGAACCTGCTGCCAAGCAACATT 12320
QY 121 tataactgttagaacactgtcttattgttggtagacatttccacaactgttata 180
|||||
DB 12321 TATTAACTGTTAACACACTTGTATTGTTGTGTGACATATTTTCCAAATGTTATA 12380
QY 181 attatataagtgtagtgaacagatgcaatcttctgtgtctaaagtgctcagttaa 240
|||||
DB 12381 ATTATATATAGTGTGTGAACAGATGCAATCTTTGTGTCTAAAGTGCTGCAGTTAA 12440
QY 241 aaaaaaaacaactttcttccaatagatgcatgtagtgaggtttttaaacttaaaac 300
|||||
DB 12441 AAAAAAAACACCTTTCTTCAATATGCGATGTAGGAGGTTTCTTAACTTTAAATAC 12500
QY 301 atcaaaaatgttaaaaatcaatctgttatctagtagttaattatcgcttatattc 360
|||||
DB 12501 ATCAAAAATGTGTTAAATCATCTGTGTATCTAGTAGTTTAAATATCGCCTTATATTC 12560
QY 361 cccatgaatgaatgaacactgaattaatcatggttgcctgcacatgcttcttactt 420
|||||
DB 12561 CCCATGATGATGACAGACTGACATTTAATTCATGTTGTGTGCGCATGCTTCTTACTTT 12620
QY 421 aacatattcttttgcgaatgtaaaagtgtaagataatagttatataagtgctcg 480
|||||

Db 12621 AACATATTTCTTTCGAGATGTAAGATATAGTTATATATAAGTACTGCG 12680
Qy 481 ctgtaaatgctgaataataactttatgcaatgaaggtctacgaacatgttgaaactt 540
Db 12681 CTGTAATGATGCTAAATTAATCTTATGCAATTAAGGGCTTACGAACATGTAACCTT 12740
Qy 541 tttctatcttacttggaataagaatgctttgaccccccacattatcttgcct 593
Db 12741 TTTTACTTTTATTTGGCAATAGGAATGTTTGGCACCCTCCACATTTTATGCTT 12793

RESULT 3
AF193556 12793 bp DNA linear PRI 07-FEB-2000
LOCUS Homo sapiens saccin (SACS) gene, complete cds.
ACCESSION AF193556
VERSION AF193556.1 GI:6907041
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 12793)
Engert,J.C., Berube,P., Mercier,J., Dore,C., Lepage,P., Ge,B.,
Bouchard,J.P., Mathieu,J., Melancon,S.B., Schalling,M.,
Lander,E.S., Morgan,K., Hudson,T.J. and Richer,A.
ARSACS, a spastic ataxia common in northeastern Quebec, is caused
by mutations in a new gene encoding an 11.5-Kb ORF
Nat. Genet. 24 (2), 120-125 (2000)
2 (bases 1 to 12793)
Engert,J.C., Berube,P., Dore,C., Lepage,P., Ge,B., Hudson,T.J. and
Richer,A.
Direct Submission
Submitted (08-OCT-1999) Genome Centre, Montreal General Hospital,
1650 Cedar Ave., Montreal, QC H3G 1A4, Canada

TITLE JOURNAL
MEDLINE
AUTHORS
FEATURES
source
1. 12793
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="between D13S232 and D13S292"
1. 12793
/gene="SACS"
/product="saccin"
1. 12793
/gene="SACS"
77. 11566
/gene="SACS"
/note="molecular chaperone"
/codon_start=1
/product="saccin"
/protein_id="AF31262.1"
/db_xref="GI:6907042"
/translation="MNTFWGRELIVQWYFEDNRHPSYMLKMYKNIYHFSFDL
TIFDEMRPLRTLEBQVCVELIRIRISLVILDDSEKQPEFLADIVOKAGFVL
KUDASIQPLAKTIHSPSPSAVIDIMEMLQKLCNQTSLTPHKAIRFLASL
TDSSEKERKIQLALFKRIHNSDQGISYTKKCKVLAHNAKLPADRLISVID
SSDEATIRLANMKIEQLTKTSLCLAVLDIENAFSHEWTEQMLMVENLSLNE
NNPVEMLTLPKLEIIOSEOMYAGELFQPDIVLDFCNEGTEFPVPFSPDIL
HSLROGLKNEASLKEIKEDVVOYAKTEALVOYACGPDQVILKAKKLLIYANKNHL
OSSEKMTLAKTIKVPACKERPNYRPGSLWKDCLNLCAPPKCVGAHILIGSSLP
LVESTHVNLEKALGITKPSLSAVLNHFKTIVDVYSKTFSDVDYQFQHILEITGF
MDHLNEGDSFRALKFPMWWTGKFCPLAQAIVKPIHDLADQPYLHNVKMAKPHQ
LFKVCSEIELTSDHISIMYIOLYILKSDDLSEBOKNHLMLNLIHMLYSQIAPAS
PNTVPYIHHSKNPSKLMKPIHECCYCDIKVDLDLDESVPIILHVDIMPKTAE
MLKVPCLSTRLLINPEMNGFEGOSREPLRYRIKNIIEEYPSVDIKETLIONMDANA
TECSEFLDMRRMNDIRENLDPGMAACHRPALMSFNNSSQSDSDPNNITRIGESLARK
EVDKVGKFLGNSVYHTIDIPILMSREFNIMDPININHSKIKKSNGLIKINMSK
QOKRLKRPNQEPFLIDVFGQLPLVEAPYSVNGTFLRSLFTQDEAKSEVSTCY
MNAIDYSLVDESLGCHRLIFPOSVSKMLYKLTETNPSLAODTIVILKKKSSCK
ALNTPLVSLKEAKKLMKTCSSSNKPLPSDEPKSCILQITVEEFHVFRRIDLOSP
LFKGPDDDDPAALFEMAKSGSKKPSDELISQKTYECTTWLCTCMTDGEALKFLSLSSG

RLGLVPCGAVGVOLSEIOPDKMTVKRPHIGVEVCYPLRLIKTGLPYHINCFVNTSNR
KEIMKTDTKGRNMTFMRHYIVAYAYLOVLSVLDLATSGLMDTYTAAVAPDDDLVD
DESVICQGFEDIAHGKRGKLTIVFSDGSGTSMKVNRPILDDSIKRDVGSNAKFI
LKYLRKTSNKLAVELPSSVKLGFEEAGCKQLILENTSEKQFSEVFEPNLOEIA
ELRDPMLIPIVLEKVDSEFSGVLPDIPCSLGHPLVPSRLIHGGRVAKLPJDK
GRPPSYTODYLNPIILIKLVQGMKADLTIMDDMLERAVSAIVEIKNSDVAALCSS
IILSLDEKIKRDPRAKDFAAKQYOTRFLPFLPKRPGFSLDMKNGSKRETEMAAT
LYAEHODIYCLLOPLINENSHSFRGCGSILAVKEFLGKPKYDVLVNOLEKAV
SVDDGTLTIOENITNACTYLHEALMONEITKASIDKLPPSFLVENAVYSEKVS
FLNFEAPALYOLPNKYNKNNFELFETVGVQSVVEDALVESIDQGRKQOLITE
ENFOLCRRIISGCIWSLIREKQFCEKNYKILIPDTNMLIPASICYNDCPWIK
KOTTYKCHADIPREYAVKLGAVPKRKHLEASVNCFTTLTEFOKRLSRIS
ILNAYPSEKMEKLELONADAKATEICFEPDRORHVDPIDDKAPALOGPRLCYVN
NOPPTEDVRCIGNLKGTEGPNPKYQGIQGFNSYHTITDDEPSITSCDICTDP
HANYAPGATISIPGRMRDIDADRFQFSLVDLYLGTFRKLDCTMRPRLNAEWA
KVEISVSPASDMVONLIDKLRSDGAEELMPLNHEKISICEIDSGTGLNVLVSK
GKLTIDGRLKROPHASVIDSVTKRROLKDIPIVOYITTYMDTEBSEGNLTWMLJONS
GFSSMEKVSYSIAHKNODITLLEPRGVAACTIHNYKPHRAKCFPLPLEGLP
VNGHPLDASARRMLMRDNGVGRSDNNLSLMTALAPAVVELIOLKRLYPGSDPT
LSYLONTPRHVAVDTLKKLSPFPVNRDLQDPLXCLVAKLYNCHEDMRLLPYVA
PNTGSDLSHVAITITWIMSTSKTRFFPNLQDELQHLKMDYNTTKRYAENVY
RLKHLLEIGFNLVYNCDETANLYHCLIDADIVSVTPADISFLMTFESSPPTNCH
GKLPCRLOQTNLKLHFLSKLVDVCFEDAEENEIEVGLPLITLDSVLOTFPARKRK
FLTYHELIPSRDLEFNTLYLKYSNLNCYKAKVPISSPDLSSVLPREKTKS
CTKKMDPASESLKNAWHFISSESVYKEOEDEFKPPDIVDTLKMALPLPGKFTV
SANOLVYBESDVLPLSLMIATVPNASSKVFHALKACCIOLALNKTSKSAVP
LISCHTANISPTSLKALHMYOTSTFRAEKLEVENDEFLMIFNCLNHLMSODI
KILKSLPCYSISIGRAYSIGKFGCYVLLTSPISAPEKQWSSSAFLEKILHLE
YEVIGIPVDDLEKYLHLLKLENI.SYDALKLEHLIYLKRLSABEELSIKLEQLE
LESLLIHDANSRLKOKHFDYDRVREYVEPEKLEIPDFEKLKOLIKPNHVF
MTSWERLRNIGLKYILSOOLLOFAKEISVANTENWSEKTLONTNYDILNHPOR
MDLSGFLKELSLIPLCERBAPEIRRHPOYQENVGLPLIKFNGAIVPARKOC
DVQLMTSCPIPERKATPLSIKEBESDLPBQOLVNLMLVNDPLVDVYVINC
RNICNTYTLDEEAWKTRAKVLRSLYEFLSEKKEFRQLGVAFVAMDEKFLKPE
VVNLIEYSDPKFYLYKLPLELTFHOLFHELTEDIISFQYVEVYSRIFKSEGO
LDENKRTYRVVSGLEFRSLONDSVYKRSLOVDLALPDSODGVSLSILFEDE
APHYKRTIGNIGVOMLYDSOCYLGKDHGFHKMLPPOKLRPLLSLILEQDE
ETPRKCOFGALCSLOGRQLDLSSEQITTLIRIKMEENDNAPLANEKARILCKALR
EGLKVSCEFLQTTLRVKGFPNIPHSSEFPAFLKRGVNLILLYIQHSDKIDNPL
ALAMTKLSANDNLISDYSILAMIGCNDIRIEBLKRDGLVCKYSSKLELPMPT
PIPAEILYTLMDPMNVYFGEVYGVLDAGGDIVYSYOPTYALIIYQVEVEDAD
NSSFLKTYIOLIDISEYKIVSSLDLYKFRPSESSRSRASPSTSPLEFLPGAR
SIRPLPSGRSHKTSKSHQSPKTKVNSLEILLKEYSVYQWAKLPDESRRKTIIRL
YLKHPDKNPENDIANEVKRLHONELNEKCPSPAGQYVSRFPVPPFKSVGNVER
KYSFQRPYTSWNOEATSHKSEROQNNKPKPSAGQYVSRFPVPPFKSVGNVER
KMLROARANFSAAARNDLHKNNANWCVCKSLSTRLAIADYAVRGSDDVPTALA
OKLIEYSQOLEGLTNDVHTLEAVGVDSLKRYPDLPLFPQIPMDRFTSVSAVMRECT
ACIIIKLENMOQKV"

BASE COUNT * 4163 a 2256 c 2487 g 3687 t
ORIGIN

Query Match 100.0%; Score 593; DB 9; Length 12793;
Best Local Similarity 100.0%; Pred. No. 3 6e-92;
Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 acatctatgtttacaggtctctcgtttgatgaagatagcaacggaataactcaaatgct 60
Db 12201 ACATCTTATGTTTACAGGCTCTCGTTTGTGAGAGATAGCAACGGAATACTCAAAATGT 12260
Qy 61 ggcagttcttatccagtggttgatgttttttgcgaaactgcttgccaagaacactt 120
Db 12261 GGCAGTCTTATVACCACTGTGTAGTATGTTTTCGAAACTGCTTCGCAACACACTT 12320
Qy 121 tattaactgttagaaccttgccttatgttgctgtcaacatctttccacaatgttca 180
Db 12321 TATTAACTGTAGAAACTTCTTATGTTGTGTGTAARATTTTCCACAAATGTTATA 12380
Qy 181 attataagtggttggaacagatgaactcttcttgctcctaaggctgctgaagtaa 240
Db 12381 ATTATATATAGTGTGTTGAACAGATCAATCTTTTGTGCTTAAAGTCTCGCAGTTAA 12440
Qy 241 aaaaaaaaaaaccttctcttcaatatgagcatgtagtgaggtctttcttaacttcaaaaac 300

```

|||||
Db 12441 AAAAAAAAAACCTTTCTTCAATATGCGATGAGCGATTGTTTAACTTAAAAAC 12500
QY 301 atcaaaatgtaaaatcattgtaataagtaataatcgttatattc 360
Db 12501 ATCAAAATGTTAAATCAATGTTATCTAGTATTAATATCGCGCTATATTTTC 12560
QY 361 cccatgaatgatcagaactacattcaatcagttgtctgcacagtccttactt 420
Db 12561 CCCATGAATGATCAGAACTGACATTAATCATGTTTGCTCGCAGCTCTTTACTTT 12620
QY 421 aacatattccttgcagaatgtaaaagtaagataaattagttatataagtgactcg 480
Db 12621 AACATATTTCTTTTGCAGAAATGTAAGATGATTAATTAATATGCTACTCG 12680
QY 481 ctgtaaatgctgaataacttacttctgcaataaggctctacagaaacagttgaact 540
Db 12681 CTGTAATGATGCTAAATATCTTTATGCAATTAAGGCTTACAGAACTGTAACCTT 12740
QY 541 ttcttacttattgggaataagaatgttgcacctccacatttacttct 593
Db 12741 TTTTACTTTATTTGGAATAAGAAATGTTTGACCTCCACATTTATTTCTT 12793

RESULT 4
AL157766 92693 bp DNA linear PRI 11-APR-2001
LOCUS AL157766/c
DEFINITION Human DNA sequence from clone RP11-40020 on chromosome
ACCESSION AL157766
VERSION AL157766
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 92693)
AUTHORS Tromans, A.
TITLE Direct Submissiion
JOURNAL Submitted (11-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT requests: clone@sanger.ac.uk
On Apr 12, 2001 this sequence version replaced gi:12709868.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/Celegans/WormPEP
This sequence
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
RP11-40020 is from the library RPc1-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-40020. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP11-760M1 is at 92594 in this sequence.

```

```

FEATURES             The true right end of clone RP11-72P19 is at 100 in this sequence.
source               Location/Qualifiers
                    1..92693
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="13"
                    /map="q12.11-12.2"
                    /clone="RP11-40020"
                    /clone_1lb="RPc1-11.1"
                    2390..2485
                    /note="MIR repeat: matches 81..192 of consensus"
                    2562..2673
                    /note="MIR repeat: matches 68..183 of consensus"
                    3896..4201
                    /note="AluY repeat: matches 3..308 of consensus"
                    5122..5397
                    /note="MER46C repeat: matches 1..286 of consensus"
                    18986..19294
                    /note="AluY repeat: matches 1..310 of consensus"
                    19644..19873
                    /note="MER46A repeat: matches 1..235 of consensus"
                    20613..20912
                    /note="AluSg1 repeat: matches 1..300 of consensus"
                    23342..23651
                    /note="AluSg1 repeat: matches 1..309 of consensus"
                    24769..24891
                    /note="L2 repeat: matches 2554..2682 of consensus"
                    25871..26011
                    /note="L2 repeat: matches 2356..2495 of consensus"
                    26033..26109
                    /note="L2 repeat: matches 2601..2688 of consensus"
                    26245..26344
                    /note="L2 repeat: matches 2154..2255 of consensus"
                    26938..27096
                    /note="MIR repeat: matches 3..175 of consensus"
                    27150..27653
                    /note="L2 repeat: matches 1063..1644 of consensus"
                    28522..28891
                    /note="THE1B repeat: matches 1..364 of consensus"
                    29447..29834
                    /note="LIME3A repeat: matches 5787..6164 of consensus"
                    36098..36415
                    /note="AluSx repeat: matches 1..308 of consensus"
                    37202..37414
                    /note="MIR repeat: matches 22..262 of consensus"
                    37963..38254
                    /note="AluSg repeat: matches 9..301 of consensus"
                    38703..39008
                    /note="AluSg repeat: matches 1..306 of consensus"
                    39790..40093
                    /note="AluSx repeat: matches 1..304 of consensus"
                    40126..40416
                    /note="AluSg repeat: matches 1..292 of consensus"
                    40444..40733
                    /note="AluSg repeat: matches 1..292 of consensus"
                    41322..41405
                    /note="Single clone region. Assembly confirmed by
                    restriction digest data."
                    41541..41788
                    /note="AluSg repeat: matches 1..248 of consensus"
                    44790..45101
                    /note="AluSg repeat: matches 1..313 of consensus"
                    45261..45312
                    /note="13 copies 4 mer tggc 888 conserved"
                    45899..46206
                    /note="AluY repeat: matches 1..307 of consensus"
                    46754..47052
                    /note="AluY repeat: matches 1..298 of consensus"
                    47067..47365
                    /note="AluY repeat: matches 1..299 of consensus"
                    47477..47873
                    /note="L1MA10 repeat: matches 5950..6322 of consensus"
                    47889..48229
                    repeat_region

```

```

repeat_region /note="AluX repeat: matches 1. .312 of consensus"
49168. 49212
repeat_region /note="Alu repeat: matches 85. .126 of consensus"
49620. .49693
repeat_region /note="L2 repeat: matches 1685. .1757 of consensus"
50704. .51032
misc_feature /note="AluX repeat: matches 1. .308 of consensus"
52204. .53009
/note="CPG Island"
evidence=not_experimental
repeat_region /note="AluX repeat: matches 1. .160 of consensus"
53978. .54137
/note="LIMB6 repeat: matches 5822. .6172 of consensus"
54179. .54511
55685. .55949
repeat_region /note="AluX repeat: matches 1. .305 of consensus"
57331. .57390
/note="30 copies 2 mer ga 75% conserved"
57357. .57392
/note="9 copies 4 mer gaga 91% conserved"
57753. .57930
repeat_region /note="MIR repeat: matches 82. .262 of consensus"
58260. .58389
/note="MIR repeat: matches 2. .153 of consensus"
58564. .58611
repeat_region /note="24 copies 2 mer ca 93% conserved"
59350. .59533
/note="AluX repeat: matches 129. .313 of consensus"
59922. .60223
/note="AluX repeat: matches 85. .299 of consensus"
61036. .61144
/note="L2 repeat: matches 2581. .2656 of consensus"
62008. .62187
/note="TIGER1 repeat: matches 2238. .2418 of consensus"
62188. .62316
repeat_region /note="Alu repeat: matches 1. .129 of consensus"
62330. .62363
/note="Alu repeat: matches 261. .294 of consensus"
62362. .62565
/note="TIGER1 repeat: matches 1586. .1787 of consensus"
62566. .62865
/note="AluX repeat: matches 1. .302 of consensus"
62866. .64385
/note="TIGER1 repeat: matches 46. .1586 of consensus"
64386. .64694
/note="Alu repeat: matches 1. .306 of consensus"
64695. .64713
/note="TIGER1 repeat: matches 29. .46 of consensus"
65068. .65395
/note="LIPB2 repeat: matches 5405. .5733 of consensus"
65396. .65569
/note="Alu repeat: matches 136. .309 of consensus"
65571. .65640
/note="LIPB2 repeat: matches 5728. .5791 of consensus"
65696. .65717
/note="11 copies 2 mer ta 100% conserved"
65723. .66096
/note="LIPB2 repeat: matches 5789. .6155 of consensus"
66371. .66410
/note="10 copies 4 mer tctg 82% conserved"
67586. .67886
/note="AluB repeat: matches 1. .299 of consensus"
69748. .69930
/note="MIR repeat: matches 6. .248 of consensus"
70957. .71267
/note="Alu repeat: matches 1. .311 of consensus"
71279. .71413
/note="MER2B repeat: matches 548. .680 of consensus"
71411. .71737
/note="MER3A repeat: matches 47. .485 of consensus"
71780. .72075
/note="AluX repeat: matches 1. .295 of consensus"
72145. .72256

```

```

repeat_region /note="MER31-internal repeat: matches 42. .175 of
72454. .72865
consensus"
repeat_region /note="MER31-internal repeat: matches 332. .739 of
72873. .73249
consensus"
repeat_region /note="MER31-internal repeat: matches 883. .1261 of
72873. .73249
consensus"

Query Match 100.0%; Score 593; DB 9; Length 92693;
Best Local Similarity 100.0%; Pred. No. 2.2e-92;
Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aacatctatggttaacagcttcctcgtttgtaagaataacgaagaactcaaatgct 60
DB 6709 ACATCTTATGTTTACAGGCTTCCTGTTGATGAAGATGCAACGAAACTCAAAATGCT 6650
QY 61 ggcagctcttataccagcttgtagtattgtttcgtgaagctgttccaaacacat 120
DB 6649 GCGAGTCTTATTTACAGGCTTGTAGTATGTTTCTGGAACCTGCTCCAGACACAT 6530
QY 121 tattaactgtagaacactgtcttaagtgtgtgtacataattccacaagtata 180
DB 6589 TATTACTGTAGAACACTTCTTATGTTGTGTACATATTTCCACAAATGTTATTA 6530
QY 181 attatatagttggttgaacagatgcaactctttgtgtcctaagggtgtcagttaa 240
DB 6529 ATTATATAGTGTGTGACAGATGCAATCTTTGTGTCTAAGGTGTGCGAGTTAA 6470
QY 241 aaaaaaacaaccccttcttcaatatagcatagtatgtaatttcaacttaaaac 300
DB 6469 AAAAAAACAACCTTTCTTCAATATGCAATGCAATGCAATGCAATGCAATGCAAT 6410
QY 301 atcaaaaatgttaaaatcaatcattgttatacagtaattataatatacgtctata 360
DB 6409 ATCAAAAATGTTAAATCATGTTATGTTATGTTATGTTATGTTATGTTATGTT 6350
QY 361 cccatgaatgtagaactgacatttaattcattgttcgcgaatgctcttaactt 420
DB 6349 CCGATGAATGATGCAAGATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 6290
QY 421 aacatctcttctgcagaatgtaaaagtaataatagttatataagttactg 480
DB 6289 AACATATTTCTTTGCAAGATGTAAGATGATATGTTATGTTATGTTATGTTATG 6230
QY 481 ctgtaaatgatacgaataatacttatacgaatgaaggctacgaacatgtgaact 540
DB 6229 CTGTAATGATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 6170
QY 541 tttaacttattggaataagaatgttgcacccacatttatgctt 593
DB 6169 TTTTACTTTTATGGAATAGGAATGTTGCACTCCACATTTTATGCTT 6117

RESULT 5
AC079761 99819 bp DNA linear HTG 10-SEP-2000
LOCUS AC079761
DEFINITION Homo sapiens chromosome UNK clone R11-143617, *** SEQUENCING IN
PROGRESS ***, 44 unordered pieces.
ACCESSION AC079761
VERSION AC079761.1 GI:10047966
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 99819)
AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 99819)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-2000) Genome Sequencing Center, Washington

```

COMMENT

University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information -----

```

NOTE: This is a 'working draft' sequence. It currently consists of 44 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number WILLIM11 be preserved.

*	1	1137	contig of 1137 bp in length
*	1138	1237	gap of unknown length
*	1238	2538	contig of 1301 bp in length
*	2539	2688	gap of unknown length
*	2639	3376	contig of 1338 bp in length
*	3977	4076	gap of unknown length
*	4077	5355	contig of 1279 bp in length
*	5356	5455	gap of unknown length
*	5456	6757	contig of 1302 bp in length
*	6758	6857	gap of unknown length
*	6858	8570	contig of 1713 bp in length
*	8571	8670	gap of unknown length
*	8671	9925	contig of 1255 bp in length
*	9926	10025	gap of unknown length
*	10026	11326	contig of 1401 bp in length
*	11427	11526	gap of unknown length
*	11527	13366	contig of 1740 bp in length
*	13267	13366	gap of unknown length
*	13367	14794	contig of 1428 bp in length
*	14795	14894	gap of unknown length
*	14895	16054	contig of 1160 bp in length
*	16055	16154	gap of unknown length
*	16155	17395	contig of 1241 bp in length
*	17396	17495	gap of unknown length
*	17496	19387	contig of 1792 bp in length
*	19388	19387	gap of unknown length
*	19388	21394	contig of 1907 bp in length
*	21295	21394	gap of unknown length
*	21395	22044	contig of 1550 bp in length
*	22945	23044	gap of unknown length
*	23045	24421	contig of 1377 bp in length
*	24422	24521	gap of unknown length
*	24522	25870	contig of 1349 bp in length
*	25871	25970	gap of unknown length
*	25971	27230	contig of 1260 bp in length
*	27231	27330	gap of unknown length
*	27331	28778	contig of 1448 bp in length
*	28779	28878	gap of unknown length
*	28879	30893	contig of 2015 bp in length
*	30894	30993	gap of unknown length
*	30994	32460	contig of 1467 bp in length
*	32461	32560	gap of unknown length
*	32561	33584	contig of 1424 bp in length
*	33585	34084	gap of unknown length
*	34085	35385	contig of 1201 bp in length
*	35386	35385	gap of unknown length
*	35386	37184	contig of 1799 bp in length
*	37185	37284	gap of unknown length
*	37285	39172	contig of 1868 bp in length
*	39173	39272	gap of unknown length
*	39273	40874	contig of 1602 bp in length
*	40875	40974	gap of unknown length
*	40975	42893	contig of 1919 bp in length
*	42894	42893	gap of unknown length
*	42994	44384	contig of 1391 bp in length
*	44385	44584	gap of unknown length
*	44485	44599	contig of 1515 bp in length

	*	46000	46099	gap of unknown length
	*	46100	48669	contig of 2570 bp in length
	*	46670	48769	gap of unknown length
	*	48770	50798	contig of 2029 bp in length
	*	50799	50898	gap of unknown length
	*	50899	52809	contig of 1911 bp in length
	*	52810	52909	gap of unknown length
	*	52910	55127	contig of 2218 bp in length
	*	55128	55227	gap of unknown length
	*	55228	58087	contig of 2860 bp in length
	*	58088	58187	gap of unknown length
	*	58188	61004	contig of 2817 bp in length
	*	61005	61104	gap of unknown length
	*	61105	64185	contig of 3081 bp in length
	*	64186	64285	gap of unknown length
	*	64286	67105	contig of 2820 bp in length
	*	67106	67205	gap of unknown length
	*	67206	70837	contig of 3632 bp in length
	*	70838	70937	gap of unknown length
	*	70938	75837	contig of 4900 bp in length
	*	75838	75937	gap of unknown length
	*	75938	80452	contig of 4515 bp in length
	*	80453	80552	gap of unknown length
	*	80553	84561	contig of 4109 bp in length
	*	84662	84761	gap of unknown length
	*	84762	90542	contig of 5781 bp in length
	*	90543	90642	gap of unknown length
	*	90643	94348	contig of 3706 bp in length
	*	94349	94448	gap of unknown length
	*	94449	99819	contig of 5371 bp in length
FEATURES	Location/Qualifiers			
	source	1..99819		
		/organism="Homo sapiens"		
		/db_xref="taxon:9606"		
		/chromosome="UNK"		
		/clone="RP11-143G17"		
		1..1137		
misc_feature		/note="assembly_name:Contig20"		
misc_feature		1238..2538		
misc_feature		/note="assembly_name:Contig23"		
misc_feature		2639..3976		
misc_feature		/note="assembly_name:Contig30"		
misc_feature		4077..5355		
misc_feature		/note="assembly_name:Contig35"		
misc_feature		5456..6757		
misc_feature		/note="assembly_name:Contig39"		
misc_feature		6858..8570		
misc_feature		/note="assembly_name:Contig40"		
misc_feature		8671..9925		
misc_feature		/note="assembly_name:Contig41"		
misc_feature		10026..11426		
misc_feature		/note="assembly_name:Contig42"		
misc_feature		11527..13266		
misc_feature		/note="assembly_name:Contig44"		
misc_feature		13367..14794		
misc_feature		/note="assembly_name:Contig47"		
misc_feature		14895..16054		
misc_feature		/note="assembly_name:Contig48"		
misc_feature		16155..17395		
misc_feature		/note="assembly_name:Contig49"		
misc_feature		17496..19287		
misc_feature		/note="assembly_name:Contig50"		
misc_feature		19388..21294		
misc_feature		/note="assembly_name:Contig51"		
misc_feature		21395..22944		
misc_feature		/note="assembly_name:Contig52"		
misc_feature		23045..24421		
misc_feature		/note="assembly_name:Contig53"		
misc_feature		24522..25870		
misc_feature		/note="assembly_name:Contig54"		
misc_feature		25971..27230		
misc_feature		/note="assembly_name:Contig55"		
misc_feature		27331..28778		


```

QY 552 attgggaataagaatcgttgcacccacattatgctt 593
DB 118 ATTGGGAATRAAGAAATGTTGCACCTCCACATTTATGCTT 77

RESULT 8
AC004688 169794 bp DNA linear HTG 12-AUG-2000
LOCUS Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN
DEFINITION PROGRESS ***, 4 unordered pieces.
ACCESSION AC004688.8 GI:9797726
VERSION AC004688
KEYWORDS HTG: HTGS_PHASE1.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS 1 (bases 1 to 169794)
Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J., Tamaki,T.,
Kurd,I.O.B., Conway,A.B. and Davis,R.W.
TITLE Plasmodium falciparum 3D7 chromosome 12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 169794)
Hyman,R.W., Fung,E.L., Qin,F., Conway,A.B. and Davis,R.W.
AUTHORS Direct Submission
TITLE Submitted (16-MAY-1998) Stanford DNA Sequencing and Technology
JOURNAL Center, Stanford University, 855 California Avenue, Palo Alto, CA
94304, USA
COMMENT On Aug 12, 2000 this sequence version replaced gi:8810452.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 43829: contig of 43829 bp in length
* 43830 44029: gap of unknown length
* 44030 87202: contig of 43173 bp in length
* 87203 87402: gap of unknown length
* 87403 94824: contig of 7422 bp in length
* 94825 95024: gap of unknown length
* 95025 169794: contig of 74770 bp in length.
FEATURES
source
1..169794
Location/Qualifiers
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="12"
/clone="PFYAC69"
/clone="3D7"

BASE COUNT 69682 a 15349 c 14920 g 69043 t 600 others
ORIGIN

Query Match 10.3%; Score 61; DB 2; Length 169794;
Best Local Similarity 47.9%; Pred. No. 0.063;
Matches 175; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

QY 70 tattacagtgtagtattgttctcgaaacgctgcgaagaacatttaactg 129
DB 28288 TATTAACCTGTATATATTTCTAACAATAGTACAGCTCTCAATATATATATGTCAAAGT 28347
QY 130 ttgaacactgcttattgttgcgtggtacataattccacaatgltataattatata 189
DB 28348 TTGGAAATATTTAAATATATCTCTTTTAAATGTTTAAATAATTTATTTATTTCTCTT 28407
QY 190 gtgtggtgtaacggatgcacatcttctgtctctaaagtgctgcagcttaaaaaaac 249
DB 28408 AAAAATTTATTCATGTCATATTTTAAAGTATATATATATATATATATATATATA 28467
QY 250 aaccttcttcaatacgcatcgtgagcttcttcaacttaaaacatcaaat 309
DB 28468 TATATATATATGGAATATAATATCTTTCATATTTTAAATCTTTTATATATATAT 28527

```

```

QY 310 tgttaaacatcgtgtactagtagttataataatcgtctatattcccaatgaat 369
DB 28528 ATATATATATATATTTATGATATTTTCTGCTTTTCTGCTTTTATTTATTTTATTTT 28587
QY 370 gatcagaactgacatttaactatgctgtcgcacatgcttcttaacttaacattt 429
DB 28588 TACCTATATTTATTTCTTTTTCATGATTTGCTTCATATTTTCTTTTCCATTTT 28647
QY 430 ctttt 434
DB 28648 TTTT 28652

RESULT 9
AC004709 196149 bp DNA linear HTG 01-APR-1999
LOCUS Plasmodium falciparum chromosome 12, *** SEQUENCING IN PROGRESS
DEFINITION *** 3 unordered pieces.
ACCESSION AC004709.3 GI:4558585
VERSION AC004709
KEYWORDS HTG: HTGS_PHASE1.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS 1 (bases 1 to 196149)
Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurd,I.O.B., Conway,A.B.
and Davis,R.W.
TITLE Plasmodium falciparum 3D7 chromosome 12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 196149)
Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
AUTHORS Direct Submission
TITLE Submitted (21-MAY-1998) Stanford DNA Sequencing and Technology
JOURNAL Center, Stanford University, 855 California Avenue, Palo Alto, CA
94304, USA
COMMENT On Apr 2, 1999 this sequence version replaced gi:4337173.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 47631: contig of 47631 bp in length
* 47632 47831: gap of unknown length
* 47832 179129: contig of 131298 bp in length
* 179130 179329: gap of unknown length
* 179330 196149: contig of 16820 bp in length.
FEATURES
source
1..196149
Location/Qualifiers
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="12"

BASE COUNT 80057 a 19753 c 18800 g 77138 t 401 others
ORIGIN

Query Match 10.3%; Score 61; DB 2; Length 196149;
Best Local Similarity 47.9%; Pred. No. 0.063;
Matches 175; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

QY 70 tattacagtgtagtattgttctcgaaacgctgcgaagaacatttaactg 129
DB 119329 TATTAACCTGTATATATTTCTAACAATAGTACAGCTCTCAATATATATATGTCAAAGT 119388
QY 130 ttgaacactgcttattgttgcgtggtacataattccacaatgltataattatata 189
DB 119389 TTGGAAATATTTAAATATATCTCTTTTAAATGTTTAAATAATTTATTTATTTCTCTT 119448
QY 190 gtgtggtgtaacggatgcacatcttctgtctctaaagtgctgcagcttaaaaaaac 249

```

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORFANISM
10	DMU11584/c	DMU11584	4601 bp	DNA	linear	INV 23-JUL-1994	
		Drosophila melanogaster Oregon-R mitochondrial A+T region.					
		U11584.1	GI:508826		mitochondrial DNA; A+T region; tandem repeats.		
		fruit fly.					
		Mitochondrion Drosophila melanogaster					

FEATURES	Location/Qualifiers
source	1. .4601

```

/organism="Proscophila melanogaster"
/organella="mitochondrion"
/strain="Oregon-R"
/db_xref="taxon:7227"
/dev_stage="embryo"
550..1022
/note="repeat I-A"
/rpt_type=tandem
1023..1360
/note="repeat I-B1"
/rpt_type=tandem
1361..1705
/note="repeat I-C/A"
/rpt_type=tandem
1706..2043
/note="repeat I-B2"
/rpt_type=tandem
2044..2388
/note="repeat I-C"
/rpt_type=tandem
2491..3511
/note="deoxythymidylate stretch"
3512..2648
/partial
/rpt_type=tandem
2649..3112
/note="repeat II-A"
/rpt_type=tandem

```

Query Match	10.1%;	Score 59.6;	DB 3;	Length 4601;
Best Local Similarity	45.2%;	Pred. No. 0.27;		
Matches 218;	Conservative	0;	Mismatches 264;	Indels 0;
			Gaps	0;

OY	71	attaccagttgtaattatgtttctcgggaacgcgcgtccgaacaaacttaattacgct	130
Db	665	ATTATTATTATTAATTATTTATTCGTATAATTAAATACCAATTAATTATTATTTATA	626
OY	131	tagaacacttgccttatgtcttgctgtaacatacttccacaaatgltlaacttataag	190
Db	625	TATATTATGATTTTAAATAAATTATTTATCTTTTATTTATTCATTAATAAATTTTAATA	566
OY	121	tgttgttgaacaggatgcaactcttctgtctgaagtgctgcagcttcaaaaaaaca	250
Db	565	TTTTTTTTTAATAGATTAATTTTATTTTCTGTAAAAAAATATTTTATTAATAATTT	506
OY	231	accttcttctcaatatgycatgtagtggagttcttctaacttataaaacatacaaat	310
Db	505	TCAATTTTTTTTAAAAAAAATTTTTTTTAAATTAATTATTAATTTATTTTAAAAAAT	446
OY	311	gttaaatcatgtgttatctcagtcagttataattatcgcgtctaatctcccaagatg	370
Db	445	TGTAAAAAAAAGTTTAAATACTTTAAAAAAAATTTTTTTTATGTAATTTAAAAAGAA	386
OY	371	atcagaactgacatttaattcagttgtctcgcagatgccttctaacttaacatatctc	430
Db	385	TTAATCTAATTAATTTTAAAAATTTTTTTTAAAAAAAATTTTTTTTAAAAAAAATTTT	326
OY	431	ttttgcagaaatgtaaaaggtaatgtaattagttatataagtgtaacggcgtgtaaa	490
Db	335	TTTTTAAAAAAAATTTTAAAAAAGTGAATAATTTTAAATTTTAAAAAATTA	266
OY	491	tgctaataatacttatgcaattaaaggcttcacagaacatgttgaaccttcttacttc	550
Db	265	AAATTTATTATTATTAATATTTGATTAATATTTTTTTTTGTTAAAAAATTTATTTT	206
OY	551	ta 552	
Db	205	AA 204	

RESULT 11
DMU37541/c 19517 bp DNA circular INV 02-MAR-2001
LOCUS DMU37541 19517 bp DNA circular INV 02-MAR-2001
DEFINITION Drosophila melanogaster complete mitochondrial genome.
ACCESSION U37541
VERSION U37541.1 GI:1166529
KEYWORDS
SOURCE .
ORGANISM fruit fly,
Mitochondrion Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 12511 to 12682)
Clary,D.O., Goddard,J.M., Martin,S.C., Fauron,C.M. and
Wolstenholme,D.R.
Drosophila mitochondrial DNA: a novel gene order
Nucleic Acids Res. 10 (21),6619-6637 (1982)


```

MEDLINE      83090428
REFERENCE    2 (bases 5269 to 5695)
AUTHORS      Clary,D.O., Wahlertlthner,J.A. and Wolstenholme,D.R.
TITLE        Transfer RNA genes in Drosophila mitochondrial DNA: related 5'
              flanking sequences and comparisons to mammalian mitochondrial tRNA
              genes
JOURNAL      Nucleic Acids Res. 11 (8), 2411-2425 (1983)
MEDLINE      83220794
AUTHORS      3 (bases 404 to 5272)
              de Bruijn,M.H.
              Drosophila melanogaster mitochondrial DNA, a novel organization and
              genetic code
JOURNAL      Nature 304 (5923), 234-241 (1983)
MEDLINE      83245048
AUTHORS      4 (bases 804 to 1778)
              Satta,Y., Ishiwa,H. and Chigusa,S.I.
              Analysis of nucleotide substitutions of mitochondrial DNAs in
              Drosophila melanogaster and its sibling species
JOURNAL      Mol. Biol. Evol. 4 (6), 638-650 (1987)
MEDLINE      88174373
AUTHORS      5 (bases 5268 to 13619)
              Garesse,R.
              Drosophila melanogaster mitochondrial DNA: gene organization and
              evolutionary considerations
JOURNAL      Genetics 118 (4), 649-663 (1988)
MEDLINE      88212147
AUTHORS      6 (bases 441 to 2967)
              Satta,Y. and Takahata,N.
              Evolution of Drosophila mitochondrial DNA and the history of the
              melanogaster subgroup
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 87 (24), 9558-9562 (1990)
MEDLINE      91088557
AUTHORS      7 (bases 14215 to 14512)
              Ballard,J.W., Olsen,G.J., Faith,D.P., Odgers,W.A., Rowell,D.M. and
              Atkinson,P.W.
              Evidence from 12S ribosomal RNA sequences that onychophorans are
              modified arthropods
JOURNAL      Science 258 (5086), 1345-1348 (1992)
MEDLINE      93088057
AUTHORS      8 (bases 14917 to 19517)
              Lewis,D.L., Farr,C.L., Farguham,A.L. and Kaguni,L.S.
              Sequence, organization, and evolution of the A+T region of
              Drosophila melanogaster mitochondrial DNA
JOURNAL      Mol. Biol. Evol. 11 (3), 523-538 (1994)
MEDLINE      94265822
AUTHORS      9 (bases 1 to 408; 13319 to 19517)
              Lewis,D.L., Farr,C.L. and Kaguni,L.S.
              Drosophila melanogaster mitochondrial DNA: completion of the
              nucleotide sequence and evolutionary comparisons
JOURNAL      Insect Mol. Biol. 4 (4), 263-278 (1995)
MEDLINE      96423163
AUTHORS      10 (bases 1 to 19517)
              Lewis,D.L., Farr,C.L. and Kaguni,L.S.
              Direct Submission
JOURNAL      Submitted (03-OCT-1995) Laurie S. Kaguni, Biochemistry Department,
              Michigan State University, East Lansing, MI 48824-1319, USA
FEATURES
  source
    1..19517
      /organism="Drosophila melanogaster"
      /organelle="mitochondrion"
      /db_xref="taxon:7227"
      /note="derived from new and previously submitted
      sequences; sequence is a composite containing sequences
      obtained from different Drosophila melanogaster strains"
    1..65
      /product="tRNA-Ile"
      complement(97..165)
      /product="tRNA-Gln"
    171..239
      /product="tRNA-Phe"
      /codon_start=1
      /transl_table=5
    CDS
      /product="NADH dehydrogenase subunit 2"
      /protein_id="AAC47811.1"
      /db_xref="GI:1166530"
      /translation="MFNSSSKILFITIMITGLITVTSNWGLGAMGLEINLLSFPL
      LSDNNMLMSTESLAKYLTQVLAASVLLFSSILMLKNNKNNNEINSFTSMILSL
      LKSGAAPFHEWPMMEGLTMNALMLMTWOKTAPMLKLYLILKLLISVLSYLI
      GATGGLNQLSRLKRLAFSSILGMLSSLSLMTSESTWILFFPFSLSLTPMFMIF
      KLFHLNQLSWEPYNSKILKFTLFPNLSLGGLEPPFGFLPKHLYIQQLTCLNDYFMT
      IMMSTLITLFFYLIRICYSAPMNNYFENNIMKMNNSINYNNYMTTFEFSIFGLFI
      SLEVFME"
    1264..1329
      /product="tRNA-Trp"
      complement(1322..1383)
      /product="tRNA-Cys"
      complement(1403..1468)
      /product="tRNA-Tyr"
      join(1470..1472,1474..3009)
      /note="mechanism underlying reading frame shift after
      first codon uncertain"
      /codon_start=1
      /transl_table=5
      /product="cytochrome c oxidase subunit I"
      /protein_id="AAC47812.2"
      /db_xref="GI:7412849"
      /translation="MSRQMLFSTNHMDIGTLFIFGAGAGWGTSLISILRAELHPG
      ALGDDQITNIVTAHFTIIFPMPIMIGFGNNLVPLMLGAPMAPPRMNNKSEW
      LLEPPALLSVSMVENGAGTWTVPPLSAGIAGHGSADLAIFSLHAGISLIGA
      VNEITIVIMNRSTGISLDRMPLEFVMSVVTALILLISPLVLAGATMLTDRNLNFS
      FDPAGGDPILYQHLFMPFGHPVEYLLIPFGFMISHIISOESGKETFGSGIMYAM
      IATGILGFIYMAHMFYMGMDVDTRAVFSATMIVAPVPGIKIFESMLATLHGQISYS
      PALMALGFVEFLFVYGVGLGVANSVDIILHDITYVVAHRYVLSKQAVRAIMAGF
      IHMYPLGTLLNNKMLKSHFTIIMFGVNLTFEPQHLGLAGAPRYSDTPDYVTWN
      IVSTIGTSLTGLILFFFIITWESLVSORQVIYPIQLNSSIEWYQNTPPAHSYSELP
      LVLN"
    3012..3077
      /product="tRNA-Leu"
      3083..3767
      /note="TAA stop codon is completed by the addition of 3' A
      residues to the mRNA"
      /codon_start=1
      /transl_except=(pos:3767,aa:TERM)
      /transl_table=5
      /product="cytochrome c oxidase subunit II"
      /protein_id="AAC47813.1"
      /db_xref="GI:1166532"
      /translation="MSTWANLGLQDSASPUMEQLIFPHDHALILVITVLGYLWF
      LFNNNVNRELLHGLIEMITILPAIILIFLALPELRLLYLDDEINERSVTLKSGH
      OKWWSYEYSDPNNEIDPSYMIPTNELMTGFRLLDNDNVVLPMSNPIILVYADVI
      HSTVVALYKVDGTGPRLNQTNFTINRGLFLFGQCEISGANHSMPPIVIESVPVY
      FIKWISSNNS"
    3768..3838
      /product="tRNA-Lys"
      3840..3906
      /product="tRNA-Asp"
      3907..4068
      /codon_start=1
      /transl_table=5
      /product="ATPase 8"
      /protein_id="AAC47814.1"
      /db_xref="GI:1166533"
      /translation="MPQMAPISWLLFIIFSTIFILFCSINITYSPNPKSNELKNI
      NLNSMNKW"
    4062..4736
      /codon_start=1
      /transl_table=5
      /product="ATPase 6"
      /protein_id="AAC47815.1"
      /db_xref="GI:1166534"
      /translation="MMTNLFSVFPDLAIFNFSIWLSTFLGLMLPISYIWLMPSRNYI
      MNNSILPLTHKEKFTLGGSGHNGSTFISLSFLFNNFGLPEFYIFTSTSLTLT
      LSLALPLMLCFMVLGWINHTQHFALHVOGPAIILMPNVCIEFTSNIIIRGDTLAVR
      LTANMAGHLLLTLLGNTGSSMSYMLMTFLMAQIALLVLESVAMIGSYVAVAVSTL
      YSEVY"

```

[illegible][illegible]

[illegible]

Center code: MIBR		
Web site: http://www.seq.wi.mit.edu		
Contact: sequence_submissions@genome.wi.mit.edu		
Project Information		
Center project name: L6844		
Center clone name: 661_A23		

NOTE: This record contains 67 individual sequencing reads that have not been assembled into contigs. Runs of N are used to separate the reads and the order in which they appear is completely arbitrary. Low-pass sequence sampling is useful for identifying clones that may be gene-rich and allow overlap relationships among clones to be deduced. However, it should not be assumed that this clone will be sequenced to completion. In the event that the record is updated, the accession number will be preserved.		
1	849:	contig of 849 bp in length
850	949:	gap of 100 bp
950	1760:	contig of 811 bp in length
1761	1860:	gap of 100 bp
1861	2678:	contig of 818 bp in length
2679	2778:	gap of 100 bp
2779	3603:	contig of 823 bp in length
3604	3703:	gap of 100 bp
3704	4497:	contig of 794 bp in length
4498	4597:	gap of 100 bp
4598	5410:	contig of 813 bp in length
5411	5510:	gap of 100 bp
5511	6303:	contig of 799 bp in length
6310	6409:	gap of 100 bp
6410	7184:	contig of 775 bp in length
7185	7284:	gap of 100 bp
7285	8075:	contig of 791 bp in length
8076	8175:	gap of 100 bp
8176	8987:	contig of 812 bp in length
8988	9087:	gap of 100 bp
9088	9902:	contig of 815 bp in length
9903	10002:	gap of 100 bp
10003	10822:	contig of 820 bp in length
10823	10922:	gap of 100 bp
10923	11722:	contig of 800 bp in length
11723	11822:	gap of 100 bp
11823	12582:	contig of 760 bp in length
12583	12682:	gap of 100 bp
12683	132489:	contig of 807 bp in length
132490	135889:	gap of 100 bp
13590	14355:	contig of 766 bp in length
14356	14455:	gap of 100 bp
14456	15239:	contig of 784 bp in length
15240	15339:	gap of 100 bp
15340	16151:	contig of 812 bp in length
16152	16251:	gap of 100 bp
16252	17044:	contig of 793 bp in length
17045	17144:	gap of 100 bp
17145	17948:	contig of 804 bp in length
17949	18048:	gap of 100 bp
18049	18855:	contig of 807 bp in length
18856	18935:	gap of 100 bp
18956	19755:	contig of 804 bp in length
19760	19859:	gap of 100 bp
20664	20664:	contig of 805 bp in length
20765	20764:	gap of 100 bp
21567	21567:	contig of 803 bp in length
21568	21667:	gap of 100 bp
22483	22483:	contig of 816 bp in length
22484	22583:	gap of 100 bp
23381	23380:	contig of 797 bp in length
23481	23480:	gap of 100 bp
24291	24291:	contig of 811 bp in length
24392	24391:	gap of 100 bp
25190	25190:	contig of 799 bp in length

*	25191	25290:	gap of	100 bp	
*	25291	26110:	contig of 820 bp	in length	
*	26111	26210:	gap of	100 bp	
*	26211	27023:	contig of 813 bp	in length	
*	27024	27123:	gap of	100 bp	
*	27124	27914:	contig of 791 bp	in length	
*	27915	28014:	gap of	100 bp	
*	28015	28006:	contig of 792 bp	in length	
*	28007	28906:	gap of	100 bp	
*	28907	29718:	contig of 812 bp	in length	
*	29719	29818:	gap of	100 bp	
*	29819	30609:	contig of 791 bp	in length	
*	30610	30709:	gap of	100 bp	
*	30710	31489:	contig of 780 bp	in length	
*	31490	31589:	gap of	100 bp	
*	31590	32396:	contig of 807 bp	in length	
*	32397	32496:	gap of	100 bp	
*	32497	33283:	contig of 787 bp	in length	
*	33284	33383:	gap of	100 bp	
*	33384	34154:	contig of 771 bp	in length	
*	34155	34254:	gap of	100 bp	
*	34255	35100:	contig of 846 bp	in length	
*	35101	35200:	gap of	100 bp	
*	35201	36031:	contig of 831 bp	in length	
*	36032	36131:	gap of	100 bp	
*	36132	36994:	contig of 823 bp	in length	
*	36995	37054:	gap of	100 bp	
*	37055	37959:	contig of 905 bp	in length	
*	37960	38059:	gap of	100 bp	
*	38060	38854:	contig of 795 bp	in length	
*	38855	38954:	gap of	100 bp	
*	38955	39764:	contig of 810 bp	in length	
*	39765	39864:	gap of	100 bp	
*	39865	40647:	contig of 783 bp	in length	
*	40648	40747:	gap of	100 bp	
*	40748	41559:	contig of 812 bp	in length	
*	41560	41659:	gap of	100 bp	
*	41660	42448:	contig of 789 bp	in length	
*	42449	42548:	gap of	100 bp	
*	42549	43343:	contig of 795 bp	in length	
*	43344	43444:	gap of	100 bp	
*	43444	44244:	contig of 791 bp	in length	
*	44235	44334:	gap of	100 bp	
*	44335	45113:	contig of 809 bp	in length	
*	45114	45243:	gap of	100 bp	
*	45244	46041:	contig of 798 bp	in length	
*	46042	46141:	gap of	100 bp	
*	46142	46937:	contig of 796 bp	in length	
*	46938	47037:	gap of	100 bp	
*	47038	47830:	contig of 793 bp	in length	
*	47831	47930:	gap of	100 bp	
*	47931	48774:	contig of 804 bp	in length	
*	48775	48834:	gap of	100 bp	
*	48835	49633:	contig of 799 bp	in length	
*	49634	49733:	gap of	100 bp	
*	49734	50517:	contig of 784 bp	in length	
*	50518	50617:	gap of	100 bp	
*	50618	51470:	contig of 853 bp	in length	
*	51471	51570:	gap of	100 bp	
*	51571	52367:	contig of 817 bp	in length	
*	52368	52487:	gap of	100 bp	
*	52488	53299:	contig of 812 bp	in length	
*	53300	53399:	gap of	100 bp	
*	53400	54204:	contig of 805 bp	in length	
*	54205	54304:	gap of	100 bp	
*	54305	55129:	contig of 825 bp	in length	
*	55130	55229:	gap of	100 bp	
*	55230	56063:	contig of 834 bp	in length	
*	56164	56693:	contig of 776 bp	in length	
*	56940	57039:	gap of	100 bp	
*	57040	57861:	contig of 812 bp	in length	
*	57852	57951:	gap of	100 bp	

FEATURES	source
*	57952 58782: contig of 831 bp in length
*	58783 58883: gap of 100 bp
*	58883 59712: contig of 830 bp in length
*	59713 59812: gap of 100 bp
*	59813 60604: contig of 792 bp in length
	Location/Qualifiers
	1..60604

Query Match	9.68;	Score 57.2;	DB 2;	Length 60604;
Best Local Similarity	43.18;	Pred. No. 0.37;		
Matches 185;	Conservative	0;	Mismatches 244;	Indels 0;
			Gaps	0;

QY 134 aacctgttattgtgtacatatltccacaatgttataattatatgtgt 193

Db 10327 AATAATATANTATTTTNNATTNNAAAAATTTNTTNNANTATANATATATTATNTTAAATTATTA 10380

Ov 194 atttcaaaacatccaaaattttctcctcctaacttaaaaaaacccc 253

100

[illegible]

D5 1044 / IIAIINIANIIANNAAIIIIAIIAIIIIANIIAAIAIAIAIAIAIAIIINII 10500

314 aaatcattgtgtatctagtagttataatattatcgcttataattcccatgaatgac 373

Db 10507 TAAATTTTNTNTTTTNTTTTATATATNTTATTTTNTNTTTNT 10566

374 aaactacatttaattcatatttatctcaccatactcttactttaacatatctt 433

10567 10624

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

[illegible]

494 taaatacttataagggcttacagaacatgtgaaac ttttactttat 553

Db 10687 TTTNATTTAATAATTTTATTATTAAATAATATATATATTTNTTCTAATAT 10746

QY 554 tgggaataa 562

Db 10747 TTATAATAA 10755

RESULT. 14

LOCUS	AC012204	183813 bp	DNA	linear	HTG 28-FEB-2001
-------	----------	-----------	-----	--------	-----------------

DEFINITION **HOMO sapiens** CHROMOSOME 4 CIVILE AFFAI 310013 map 4, MONKING DATE
SEQUENCE, 21 unordered pieces.

ACCESSION	AC012204
VERSION	AC012204.3
GT	GT:7107763

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

ORGANISM Homo sapiens

Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
Eucaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi

REFERENCE 1 (bases 1 to 183813)
 Authors: Dierker B, Johnston I, Nuechbaum C and Tander E

TITLE
Homo sapiens chromosome 4, clone RP11-315D13

REFERENCE 2 (bases 1 to 183813)

WOLHORS
BILLEN, B., LINCOLN, E., NUSBAUM, C., LAUNDER, E., ALLEN, N., ANDERSON,
Baldwin, J., Barna, N., Beckerly, B., Boukhalter, J., Boukhalter, J.,

Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,

Ferreira, P., FiltzHugh, W., Forrest, C., Funke, R., Gage, D.,

Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.

McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J., Lemoczky, J., Lieu, C., Locke, K., Macdonald, P., Malquis, N.,

TITLE
JOURNAL

COMMENT

Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
 Stange-Rhoman, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testaye, S., Tittel, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
 Wyman, D., Ye, W.-J., Zimmer, A. and Zody, M.
 Direct Submission

Submitted (21-OCT-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Feb 28, 2000 this sequence version replaced g1:6524238.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www.seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center Project name: L1740

Center Clone name: 315.D.13

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 136434 bases at least Q40

Consensus quality: 161213 bases at least Q30

Consensus quality: 173987 bases at least Q20

Insert size: 190000; agarose-fp

Insert size: 181813; sum-of-ctrls

Quality coverage: 3.6 in Q20 bases; agarose-fp

Quality coverage: 3.8 in Q20 bases; sum-of-ctrls

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 21 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1      1240: contig of 1240 bp in length
*
* 1241 1340: gap of 100 bp
*
* 1341 2108: contig of 768 bp in length
*
* 2109 2208: gap of 100 bp
*
* 2209 3537: contig of 1329 bp in length
*
* 3538 3637: gap of 100 bp
*
* 3638 6473: contig of 2836 bp in length
*
* 6474 6573: gap of 100 bp
*
* 6574 10752: contig of 4179 bp in length
*
* 10753 10852: gap of 100 bp
*
* 10853 15130: contig of 4278 bp in length
*
* 15131 15230: gap of 100 bp
*
* 15231 17255: contig of 2025 bp in length
*
* 17256 17355: gap of 100 bp
*
* 17356 22493: contig of 5138 bp in length
*
* 22494 22593: gap of 100 bp
*
* 22594 31215: contig of 8622 bp in length
*
* 31216 31315: gap of 100 bp
*
* 31316 36161: contig of 4846 bp in length
*
* 36162 36261: gap of 100 bp
*
* 36262 43637: contig of 7376 bp in length
*
* 43638 43737: gap of 100 bp
*
* 43738 50536: contig of 6799 bp in length
*
* 50537 50636: gap of 100 bp
*
* 50637 58712: contig of 8076 bp in length
*
* 58713 58812: gap of 100 bp
*
* 58813 68091: contig of 9279 bp in length
*
* 68092 68191: gap of 100 bp
*
* 68192 79031: contig of 10840 bp in length
*
* 79032 79131: gap of 100 bp
*
* 79132 91039: contig of 11908 bp in length
*
* 91040 91139: gap of 100 bp
*
* 91140 105647: contig of 14508 bp in length

```

```

* 105648 105747: gap of 100 bp
* 105748 121791: contig of 16044 bp in length
* 121792 121891: gap of 100 bp
* 121892 139013: contig of 1712 bp in length
* 139014 139113: gap of 100 bp
* 139114 156558: contig of 17545 bp in length
* 156559 156758: gap of 100 bp
* 156759 183813: contig of 27055 bp in length.
Location/Qualifiers
1. 183813
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-315D13"
/clone_lib="RP11 Human Male BAC"
1. 1240
/note="assembly_fragment"
1341. 2108
/note="assembly_fragment"
clone_end:T7
vector_side:right"
2209. 3537
/note="assembly_fragment"
3638. 6473
/note="assembly_fragment"
6574. 10752
/note="assembly_fragment"
10853. 15130
/note="assembly_fragment"
15231. 17255
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
17356. 22493
/note="assembly_fragment"
22594. 31215
/note="assembly_fragment"
31316. 36161
/note="assembly_fragment"
36262. 43637
/note="assembly_fragment"
43738. 50536
/note="assembly_fragment"
50637. 58712
/note="assembly_fragment"
58813. 68091
/note="assembly_fragment"
68192. 79031
/note="assembly_fragment"
79132. 91039
/note="assembly_fragment"
91140. 105647
/note="assembly_fragment"
105748. 121791
/note="assembly_fragment"
121892. 139013
/note="assembly_fragment"
139114. 156558
/note="assembly_fragment"
156759. 183813
/note="assembly_fragment"

```

```

BASE COUNT 54381 a 37506 c 35430 g 54474 t 2022 others
ORIGIN

```

```

Query Match 9.6% Score 57; DB 2; Length 183813;
Best Local Similarity 46.5%; Pred. No. 0.31;
Matches 222; Conservative 0; Mismatches 250; Indels 5; Gaps 1;

```

```

Oy 113 acacattatctacttgaagaacctgccttatgtgtgtacataattccacaa 172
Db 28742 ATAAATATATTTTATATATATATATATATATATATATATATATATTTT 28683

```

[illegible][illegible]

```

FEATURES
source
-----
Center: Washington University Genome Sequencing Center
Center code: MUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
-----
Project Information
Center project name: H_NH0425A23
Drafting center: 'WIBR'
-----
location/Qualifiers
1. 197225
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="4"
-----

```

```
BASE COUNT      57652 a 39731 c 40755 g 59087 t /clone="RP11-425A23"
ORIGIN
```

Query Match 9.6%; Score 57; DB 9; Length 197225;
 Best Local Similarity 46.5%;
 Matches 222; Conservative 0; Mismatches 5; Gaps 1;

[illegible]

Search completed: May 22, 2002, 06:59:14
Job time: 9139 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 06:45:08 ; Search time 463.88 Seconds
(without alignments)
2194.814 Million cell updates/sec

Title: US-09-693-205-7_COPY_12200_12792

Perfect score: 593
Sequence: 1 aacattatgtttacagctc.....aacctcacatttatgtctt 593

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	593	100.0	1317	22	AAS29058	CDNA encoding for
2	593	100.0	12792	22	AAH20176	Human mutated spas
3	593	100.0	12793	22	AAH20174	Human spastin nucl
4	593	100.0	12793	22	AAH20178	Human mutated spas
5	593	100.0	12793	22	AAH20179	Human mutated spas
6	593	100.0	12793	22	AAH20182	Human mutated spas
7	579.2	97.7	1387	22	AAH29132	CDNA encoding for
8	56.6	9.5	11691	24	ABL34241	Human immune syste
9	55	9.3	13038	24	ABL33274	Human immune syste

10	54.8	9.2	15387	24	ABL32184	Human immune syste
11	54.6	9.2	4468	21	AAC62126	DNA encoding a cal
12	54.6	9.2	4611	21	AAC62127	DNA encoding a cal
13	54.4	9.2	13574	21	ABL33316	Human immune syste
14	52.8	8.9	17848	22	AA545323	Chemically pretrea
15	52.8	8.9	19087	24	ABL32792	Human immune syste
16	51.6	8.7	8237	22	AA546802	Tumour suppressor
17	51.4	8.7	3442	23	ABL26886	Drosophila melanog
18	51.4	8.7	18512	24	ABL32976	Human immune syste
19	51.2	8.6	5487	24	ABL33598	Human immune syste
20	51	8.6	51	22	AA176962	Human silent SNP c
21	51	8.6	51	22	AA176964	Human silent SNP c
22	51	8.6	9652	22	ABL32908	Human immune syste
23	51	8.6	11805	24	ABL33748	Human immune syste
24	50.8	8.6	15872	22	AA546519	Tumour suppressor
25	50.6	8.5	6503	22	ABL32771	Human immune syste
26	50.6	8.5	8169	22	AA546287	Tumour suppressor
27	50.4	8.5	6294	24	ABL33054	Human immune syste
28	50.2	8.5	5424	24	ABL32854	Human immune syste
29	50.2	8.5	6038	24	ABL32065	Human immune syste
30	50.2	8.5	6101	24	AA561054	Human gene regulat
31	50	8.4	17419	22	AA545393	Chemically pretrea
32	50	8.4	17419	22	ABL33295	Human immune syste
33	50	8.4	20486	24	ABL34611	Human metastasis a
34	49.8	8.4	6112	24	ABL32489	Human immune syste
35	49.8	8.4	8423	24	ABL33407	Human immune syste
36	49.6	8.4	5357	24	ABL33546	Human immune syste
37	49.4	8.3	51	22	AA176963	Human silent SNP c
38	48.8	8.2	5888	24	ABL34457	Human metastasis a
39	48.6	8.2	6109	24	ABL32326	Human immune syste
40	48.6	8.2	18624	24	ABL33702	Human immune syste
41	48.6	8.2	18624	24	ABL33702	Human immune syste
42	48.4	8.2	7359	24	ABL33862	Human immune syste
43	48.2	8.1	6126	22	AA546573	Tumour suppressor
44	48.2	8.1	6126	24	ABL33830	Human immune syste
45	48.2	8.1	6171	24	ABL33010	Human immune syste

ALIGNMENTS

RESULT 1
AAS29058
ID AAS29058 standard; CDNA; 1317 BP.
XX
AC AAS29058;
XX
DT 21-NOV-2001 (first entry)
XX
DE CDNA encoding for human DNA-binding protein #29.
XX
XX Human: DNA-binding protein; histone; chromo domain protein;
KW chromatin organisation modifier; y-box binding protein;
KW DNA organisation; gene transcription; malignant disease;
KW autoimmune disorder; rheumatic disease; genetic abnormality;
KW infectious disease; neurological disorder; gene therapy;
KW immunomodulatory; anti-HIV; anti rheumatic; anti microbial;
cytostatic; ss.
XX
XX Homo sapiens.
OS
PN WO200155162-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01305.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.

[illegible]

20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246533.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251988.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX WPI: 2001-465557/50.
XX P-PSDB: AAU18182.
XX
XX Nucleic acid molecules encoding human secreted chromosomal binding
XX proteins, used in preventing, treating or ameliorating a disorder, e.g.,
XX Alzheimer's and Parkinson's diseases -
XX
XX Claim 4; SEQ ID No 39; 561pp; English.
XX
XX The present invention relates to the isolation of novel DNA-binding
XX proteins (AAU18154-AAU18281), and cDNA and genomic sequences encoding

CC for these proteins. DNA-binding proteins such as histones, chromo
 CC (chromatin organisation modifier) domain proteins, and y-box binding
 CC proteins may contribute to diseases resulting from aberrant DNA
 CC organisation and/or gene transcription. The sequences of the invention
 CC are useful in screening assays to identify antagonists and/or agonists
 CC that may enhance or block activities mediated by DNA-binding proteins.
 CC Blockers of DNA-binding proteins may be useful in treating disorders
 CC such as malignant diseases (e.g. cancer), autoimmune disorders
 CC (e.g. diabetes mellitus), rheumatic diseases (e.g. Rheumatoid
 CC arthritis), genetic abnormalities (e.g. cystic fibrosis), infectious
 CC diseases (e.g. HIV) and neurological disorders (e.g. Alzheimer's
 CC disease). The polynucleotide sequences of the invention may also be
 CC used in gene therapy. AAS29030-AAS29157 represent cDNA sequences
 CC encoding for novel DNA-binding proteins.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.

SQ Sequence 1317 BP: 418 A; 167 C; 243 G; 489 T; 0 other:

Query Match 100.0%; Score 593; DB 22; Length 1317;
 Best Local Similarity 100.0%; Pred. No. 7.6e-114;
 Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aacattatgttacaaggtctctgtttatgaagaatagacagcaagaaactcaaatgt 60
 Db 629 aacattatgttacaaggtctctgtttatgaagaatagacagcaagaaactcaaatgt 688
 QY 61 ggcagcttctattaccagctgttagatgtttcttctgaacagctgttccaagaacact 120
 Db 689 ggcagcttctattaccagctgttagatgtttcttctgaacagctgttccaagaacact 748
 QY 121 tattacgtgtagaacactgtcttattgttctgtgtacatatcttccaaatgttata 180
 Db 749 tattactgttgaacactgtcttattgttctgtgtacatatcttccaaatgttata 808
 QY 181 atttatatgttctgttgaacagatgcaactcttctgtcttgaagagtgctgcagttta 240
 Db 809 atttatatgttctgttgaacagatgcaactcttctgtcttgaagagtgctgcagttta 868
 QY 241 aaaaaaaacacaccttcttccaatagtcagttagtgagtttttctaacttaaaac 300
 Db 869 aaaaaaaacacaccttcttccaatagtcagttagtgagtttttctaacttaaaac 928
 QY 301 atcaaaatgtttaaataatcattgttattctagtattataataatcgcgtatatctc 360
 Db 929 atcaaaatgtttaaataatcattgttattctagtattataataatcgcgtatatctc 988
 QY 361 cccatgaatgacagacacgacattatcatgtttgtctgcgcattgtcttactctt 420
 Db 989 cccatgaatgacagacacgacattatcatgtttgtctgcgcattgtcttactctt 1048
 QY 421 aacatatctcttctcagaatgtaaaaggtaatgataatgatttatataagtgactcg 480
 Db 1049 aacatatctcttctcagaatgtaaaaggtaatgataatgatttatataagtgactcg 1108
 QY 481 ctgtaaatgagtcaataatactattatgcaatgaagggtctcagacacatgttgaact 540
 Db 1109 ctgtaaatgagtcaataatactattatgcaatgaagggtctcagacacatgttgaact 1168
 QY 541 ttttaactttatgggaataaggaatgtttgcacctcacaatttatgctt 593
 Db 1169 ttttaactttatgggaataaggaatgtttgcacctcacaatttatgctt 1221

AAH20176

AAH20176

AAH20176 standard: DNA: 12792 BP.

AAH20176;

09-AUG-2001 (first entry)

XX Human mutated spastin nucleotide sequence SEQ ID NO:7.

DE Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
 XX autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.

XX Homo sapiens.
 OS Synthetic.

FT Key Location/Qualifiers
 FT CDS 77..6604

FT /tag= a
 FT /product= "mutated spastin"

PN MO200129266-A2.

XX 26-APR-2001.

PD 20-OCT-2000; 2000MO-US29130.

PF 20-OCT-1999; 99US-0160588.

PR (UYMC-) UNIV MCGILL.

PA (HOP-) HOPITAL SAINTE-JUSTINE.

XX Hudson TJ, Engert J, Richter A;

DR WPI: 2001-308494/32.

XX P-PSDB: AAB97821.

PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PR useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -

XX Claim 1; Page -: 76pp; English.

XX The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (II) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with Spastin
 CC protein or spastin gene expression or activity. The present sequence
 CC encodes a mutated human spastin.
 CC N.B. The present sequence is not given in the present specification but
 CC is derived from the human spastin nucleotide sequence (AAH20174) as
 CC stated on page 14.

SQ Sequence 12792 BP: 4163 A; 2256 C; 2487 G; 3886 T; 0 other:

Query Match 100.0%; Score 593; DB 22; Length 12792;
 Best Local Similarity 100.0%; Pred. No. 9.7e-114;

Matches	593;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	acatcttatggttaacagcttcctctgtttgatgaagatagcaacggaacccaatggt	60						
Db	12200	acatcttatggttaacagcttcctctgtttgatgaagatagcaacggaacccaatggt	12259						
QY	61	ggcagttcttattacacagttgttagtattgtttcttggaacactgcttgccaagacaact	120						
Db	12260	ggcagttcttattacacagttgttagtattgtttcttggaacactgcttgccaagacaact	12319						
QY	121	tattatagtttgaacacactgctttatagtgttggtgtaataattttcccaaatgttata	180						
Db	12320	tattatagtttgaacacactgctttatagtgttggtgtaataattttcccaaatgttata	12379						
QY	181	attatatagttggttggaacagatgacatcttctgtgtctaaaggtgcgcagttaa	240						
Db	12380	attatatagttggttggaacagatgacatcttctgtgtctaaaggtgcgcagttaa	12439						
QY	241	aaaaaaacacaccttcttccaataatgcatgtagtggagttttttaactttaaac	300						
Db	12440	aaaaaaacacaccttcttccaataatgcatgtagtggagttttttaactttaaac	12499						
QY	301	atcaaaaattgttaaatatcatgtgttatctagtattatataatcatgcgttatattc	360						
Db	12500	atcaaaaattgttaaatatcatgtgttatctagtattatataatcatgcgttatattc	12559						
QY	361	cccatgaatgacacagactgacattatcatctgtgtcgcgcagcttcttaactt	420						
Db	12560	cccatgaatgacacagactgacattatcatctgtgtcgcgcagcttcttaactt	12619						
QY	421	aacataattcttcttcagaaatgtaaaaggtaaatgaattatataatgaatgtagt	480						
Db	12620	aacataattcttcttcagaaatgtaaaaggtaaatgaattatataatgaatgtagt	12679						
QY	481	cgttaaatgagtcttaataatacttatatgaatgaagggtttacagaactgttgaact	540						
Db	12680	cgttaaatgagtcttaataatacttatatgaatgaagggtttacagaactgttgaact	12739						
QY	541	ttttacttttatgtggaataaggaatgttgcacctccacatttatgtctt	593						
Db	12740	ttttacttttatgtggaataaggaatgttgcacctccacatttatgtctt	12792						

RESULT 3

AAH20174
ID AAH20174 standard; DNA: 12793 BP.

AC AAH20174;

DT 09-AUG-2001 (first entry)

DE Human spastin nucleotide sequence SEQ ID NO:1.

Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
reduced motor nerve velocity; hypermyelination of retinal nerve fibres;
atrophy of upper cerebellar vermis; absence of Purkinje cell;
abnormal neuronal lipid storage; genetic disorder; characterisation; ds.

OS Homo sapiens.

XX Key Location/Qualifiers

XX CDS 77..11566

XX FT /*tag= a

XX FT /product= "spastin"

XX PN WO200129266-A2.

XX PD 26-APR-2001.

XX PF 20-OCT-2000; 2000WO-US29130.

XX

PR 20-OCT-1999; 99US-0160588.

XX

PA (UYMC-) UNIV MCGILL.

PA (HOB1-) HOPITAL SAINTE-JUSTINE.

XX

PI Hudson TJ, Engert J, Richter A;

XX

DR WPI, 2001-308494/32.

XX P-PSDB; AAB97819.

PT New isolated polynucleotide, encoding spastin gene, and polypeptides,

PT useful for diagnosing autosomal recessive spastic ataxia of

PT Charlevoix-Saguenay disease by detecting two point mutations in spastin

PT gene sequence -

PS

XX Claim 1; Fig 9; 76pp; English.

The present invention describes human and mouse spastin, and mutated human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS)) gene sequences (1). The spastin gene has been mapped to chromosome 13q11. (1) have neuroprotective activities and can be used in gene therapy and as a spastin polypeptide agonists. (1), their fragments or their complements can be useful for assaying the presence of a nucleic acid molecule in a sample. (1) is useful for diagnosing or aiding in the diagnosis of an early onset neurodegenerative disease in an individual. The neurodegenerative disease comprises reduced sensory nerve conduction, reduced motor nerve velocity, hypermyelination of retinal nerve fibres, atrophy of upper cerebellar vermis, absence of Purkinje cells and abnormal neuronal lipid storage. (1) can also be used to produce antisense nucleic acids, is useful as molecular weight or chromosome markers, to identify genetic disorders, as hybridisation probes or primers, as an antigen, identify and express recombinant protein for analysis, characterisation or therapeutic use, or as markers for tissues in which the corresponding protein is expressed. Diagnostic methods from the present invention can be used to identify subjects having or at risk of developing a disease or disorder associated with aberrant expression or activity of (1). The assays can be utilised to identify a subject having or at risk of developing a disorder associated with spastin protein or spastin gene expression or activity. The present sequence encodes human spastin as given in the present invention.

Sequence 12793 BP; A163 A; 2256 C; 2487 G; 3887 T; 0 other;

Query Match 100.0%; Score 593; DB 22; Length 12793;

Best Local Similarity 100.0%; Pred. No. 9, 7e-114;

Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	acatcttatggttaacagcttcctctgtttgatgaagatagcaacggaacccaatggt	60						
Db	12201	acatcttatggttaacagcttcctctgtttgatgaagatagcaacggaacccaatggt	12260						
QY	61	ggcagttcttattacacagttgttagtattgtttcttggaacactgcttgccaagacaact	120						
Db	12261	ggcagttcttattacacagttgttagtattgtttcttggaacactgcttgccaagacaact	12320						
QY	121	tattatagtttgaacacactgctttatagtgttggtgtaataattttcccaaatgttata	180						
Db	12321	tattatagtttgaacacactgctttatagtgttggtgtaataattttcccaaatgttata	12380						
QY	181	attatatagttggttggaacagatgacatcttctgtgtctaaaggtgcgcagttaa	240						
Db	12381	attatatagttggttggaacagatgacatcttctgtgtctaaaggtgcgcagttaa	12440						
QY	241	aaaaaaacacaccttcttccaataatgcatgtagtggagttttttaactttaaac	300						
Db	12441	aaaaaaacacaccttcttccaataatgcatgtagtggagttttttaactttaaac	12500						
QY	301	atcaaaaattgttaaatatcatgtgttatctagtattatataatcatgcgttatattc	360						
Db	12501	atcaaaaattgttaaatatcatgtgttatctagtattatataatcatgcgttatattc	12560						
QY	361	cccatgaatgacacagactgacattatcatctgtgtcgcgcagcttcttaactt	420						

|||||
Db 12561 cccatgatgatcgaactgcacattatcattgttctcgcgacgtcttcttactt 12620
Qy 421 aacatattctcttgcgaaatgtaaaagtaagtaataatagttataataagttactgg 480
Db 12621 aacatattctcttgcgaaatgtaaaagtaagtaataatagttataataagttactgg 12680
Qy 481 ctgtaaatgtaagtaataatacttattatgcatttaagggtcttaacagacatgttgaactt 540
Db 12681 ctgtaaatgtaagtaataatacttattatgcatttaagggtcttaacagacatgttgaactt 12740
Qy 541 ttttactttattgggaataagaatgtttgcaccctccacatttattgctt 593
Db 12741 ttttactttattgggaataagaatgtttgcaccctccacatttattgctt 12793
RESULT 4
AAH20178
ID AAH20178 standard; DNA: 12793 BP.
XX
AC AAH20178;
XX
DT 09-AUG-2001 (first entry)
XX
DE Human mutated spastin nucleotide sequence SEQ ID NO:11.
XX
KW Human: mouse; spastin; ARSACS; chromosome 13q11; identification;
KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KW reduced motor nerve velocity; hypermyelination of retinal nerve fibres;
KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN MO200129266-A2.
XX
PD 26-APR-2001.
XX
PE 20-OCT-2000; 2000WO-US29130.
XX
PR 20-OCT-1999; 99US-0160588.
XX
PA (UYMC-) UNIV MCGILL.
PA (HOP1-) HOPITAL SAINT-JUSTINE.
XX
PI Hudson TJ, Engert J, Richter A;
XX
DR WPI; 2001-308494/32.
XX
PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -
XX
PS Claim 1; Page -: 76pp; English.
XX
CC The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
CC chromosome 13q11. (I) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (I) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for

CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (I). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with Spastin
CC protein or spastin gene expression or activity. The present sequence
CC represents a mutated human spastin gene from the present invention.
CC N.B. The present sequence is not given in the present specification but
CC is derived from the human spastin nucleotide sequence (AAH20174) as
CC stated on page 14.
XX
SO Sequence 12793 BP; 4163 A; 2255 C; 2487 G; 3888 T; 0 other;
Query Match 100.0%; Score 593; DB 22; Length 12793;
Best Local Similarity 100.0%; Pred. No. 9,7e-114;
Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 acatctatgttttacaggtctccgtttgtatgaagatagcaacgaaactccaatgtgt 60
Db 12201 acatctatgttttacaggtctccgtttgtatgaagatagcaacgaaactccaatgtgt 12260
Qy 61 ggcagttcttaccagttgttagtattgtttcttggaactgcttgcgaagacactt 120
Db 12261 ggcagttcttaccagttgttagtattgtttcttggaactgcttgcgaagacactt 12320
Qy 121 tattaactgttagaacacttgcattatgtttgtgtgatacatatttcacaaatgtata 180
Db 12321 tattaactgttagaacacttgcattatgtttgtgtgatacatatttcacaaatgtata 12380
Qy 181 attatatagtgtgttgaacaagatgcaatcctttgttgcctaagggtctgcagttaa 240
Db 12381 attatatagtgtgttgaacaagatgcaatcctttgttgcctaagggtctgcagttaa 12440
Qy 241 aaaaaaaacacctttcttccaatgcatgcatgtgagttttttaactttaaac 300
Db 12441 aaaaaaaacacctttcttccaatgcatgcatgtgagttttttaactttaaac 12500
Qy 301 atcaaaaattgttaaaatcatgttgttactagtagttataatcatcogttatattc 360
Db 12501 atcaaaaattgttaaaatcatgttgttactagtagttataatcatcogttatattc 12560
Qy 361 cccatgatgatcgaactgcacatttaactatgtttgtctgcgacgtcttcttactt 420
Db 12561 cccatgatgatcgaactgcacatttaactatgtttgtctgcgacgtcttcttactt 12620
Qy 421 aacatattcttcttgcgaatgtaaaagtaagtaataatagttataataagttactgg 480
Db 12621 aacatattcttcttgcgaatgtaaaagtaagtaataatagttataataagttactgg 12680
Qy 481 ctgtaaatgtaagtaataatacttattatgcatttaagggtcttaacagacatgttgaactt 540
Db 12681 ctgtaaatgtaagtaataatacttattatgcatttaagggtcttaacagacatgttgaactt 12740
Qy 541 ttttactttattgggaataagaatgtttgcaccctccacatttattgctt 593
Db 12741 ttttactttattgggaataagaatgtttgcaccctccacatttattgctt 12793
RESULT 5
AAH20179
ID AAH20179 standard; DNA: 12793 BP.
XX
AC AAH20179;
XX
DT 09-AUG-2001 (first entry)
XX
DE Human mutated spastin nucleotide sequence SEQ ID NO:12.
XX
KW Human: mouse; spastin; ARSACS; chromosome 13q11; identification;
KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;

KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
 OS Homo sapiens.
 OS Synthetic.
 PN WO200129266-A2.
 XX 26-APR-2001.
 PD 20-OCT-2000; 2000WO-US29130.
 PF 20-OCT-2000; 2000WO-US29130.
 XX 20-OCT-1999; 99US-0160588.
 XX (UYMC-) UNIV MCGILL.
 PA (HOP-) HOPITAL SAINTE-JUSTINE.
 PI Hudson TJ, Engert J, Richter A;
 DR WPI; 2001-308494/32.
 XX
 PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -
 XX
 PS Claim 1: Page -: 76pp: English.
 XX
 CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSAKS)) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with Spastin
 CC protein or spastin gene expression or activity. The present sequence
 CC encodes a mutated human spastin.
 CC N.B. The present sequence is not given in the present specification but
 CC is derived from the human spastin nucleotide sequence (AAH20174) as
 CC stated on page 14.
 XX
 SQ Sequence 12793 BP; 4163 A; 2257 C; 2487 G; 3886 T; 0 other;

Query Match 100.0%; Score 593; DB 22; Length 12793;
 Best Local Similarity 100.0%; Pred. No. 9.7e-114;
 Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 acatctatgtttaagagctctgtttgatgaagatagcaagaaactcaaatgt 60
 DB 12201 acatcttaagtttaacaggtctctgttcatgaagatagcaagaaactcaaatgt 12260
 OY 61 ggcagttcttataccagttgttagtatgtttctgaaactgctgccaagaacatt 120
 DB 12261 ggcagttcttataccagttgttagtatgtttctgaaactgctgccaagaacatt 12320
 OY 121 tattaactgttagaacaactgtcttatgtttgtgtacatatcttcacaagtata 180

DB 12321 tattaactgttagaacaactgtcttatgtttgtgtacatatcttcacaagtata 12380
 OY 181 attatatagttgygttgaaacagatgcaactctttgttctaaagtgctcagttaa 240
 DB 12381 attatatagttgygttgaaacagatgcaactctttgttctaaagtgctcagttaa 12440
 OY 241 aaaaaaacacacctttcttcaaatgatgatgagtgagttttttaaacttaaaac 300
 DB 12441 aaaaaaacacacctttcttcaaatgatgatgagtgagttttttaaacttaaaac 12500
 OY 301 atcaaaaaattgttaaaatcaatctgttatacttagtagtttaattatcgctataatc 360
 DB 12501 atcaaaaaattgttaaaatcaatctgttatacttagtagtttaattatcgctataatc 12560
 OY 361 cccatgaatgatcagaactacattacattacattgtgttcgcacagctctcttaatt 420
 DB 12561 cccatgaatgatcagaactacattacattacattgtgttcgcacagctctcttaatt 12620
 OY 421 aacataattcttgcagaatgtaaaagtgataatgattatataagtgactcg 480
 DB 12621 aacataattcttgcagaatgtaaaagtgataatgattatataagtgactcg 12680
 OY 481 ctgtaaatgtatgctaaataactttatgcaatgaaggcttacagaacatgtgaaact 540
 DB 12681 ctgtaaatgtatgctaaataactttatgcaatgaaggcttacagaacatgtgaaact 12740
 OY 541 ttttactttattgggaataaggaatgtttgacccctcacatttatgctt 593
 DB 12741 ttttactttattgggaataaggaatgtttgacccctcacatttatgctt 12793

RESULT 6
 AAH20182
 ID AAH20182 standard; DNA; 12793 BP.
 XX
 AC AAH20182;
 XX
 DT 09-AUG-2001 (first entry)
 XX
 DE Human mutated spastin nucleotide sequence SEQ ID NO:15.
 XX
 KW Human; mouse; spastin; ARSAKS; chromosome 13q11; identification;
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 77..11566
 FT /tag= a
 FT /product= "mutated spastin"
 XX
 PN WO200129266-A2.
 XX
 PD 26-APR-2001.
 XX
 PF 20-OCT-2000; 2000WO-US29130.
 XX
 PR 20-OCT-1999; 99US-0160588.
 XX
 PA (UYMC-) UNIV MCGILL.
 PA (HOP-) HOPITAL SAINTE-JUSTINE.
 PI Hudson TJ, Engert J, Richter A;
 DR WPI; 2001-308494/32.
 DR P-PSDB; AAB97823.
 XX
 PT New isolated polynucleotide, encoding spastin gene, and polypeptides,

PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -

XX Claim 1: Page -: 76pp; English.

CC The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSA) gene sequences (I). The spastin gene has been mapped to
CC chromosome 13q11. (I) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (II), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (II) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (I) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (I). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes a mutated human spastin.
CC N.B. The present sequence is not given in the present specification but
CC is derived from the human spastin nucleotide sequence (AAH20174) as
CC stated on page 14.

XX Sequence 12793 BP; 4163 A; 2257 C; 2487 G; 3886 T; 0 other;

Query Match 100.0%; Score 593; DB 22; Length 12793;
Best Local Similarity 100.0%; Pred. No. 9, 7e-114;
Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acatctatgtttacagagctctgtttgttgaagatagcaacggaacccaacatcgt 60
DB 12201 acatctatgtttacagagctctgtttgttgaagatagcaacggaacccaacatcgt 12260
QY 61 ggcagttcttattacacagttgttagtattcttcggaacactgtccagacaacatt 120
DB 12261 ggcagttcttattacacagttgttagtattcttcggaacactgtccagacaacatt 12320
QY 121 tattaactgttaagaacctgttattgttgtgtacacattttccacaagaattata 180
DB 12321 tattaactgttaagaacctgttattgttgtgtacacattttccacaagaattata 12380
QY 181 attatatagttgtgttgaacagatgcaactctttgttcttaagaagtgctgcagttaa 240
DB 12381 attatatagttgtgttgaacagatgcaactctttgttcttaagaagtgctgcagttaa 12440
QY 241 aaaaaaacaacctttcttcaatattgcatgcatgtagtgaagtttttaacttaaaac 300
DB 12441 aaaaaaacaacctttcttcaatattgcatgcatgtagtgaagtttttaacttaaaac 12500
QY 301 atcaaaaattgttaaaaatcttctgttattatagttattataatcagcgtataattc 360
DB 12501 atcaaaaattgttaaaaatcttctgttattatagttattataatcagcgtataattc 12560
QY 361 cccatgaatgatacagaactataatcattcattgtgtctgcacagctcttatttt 420
DB 12561 cccatgaatgatacagaactataatcattcattgtgtctgcacagctcttatttt 12620
QY 421 aacatatcttcttgcagaatgtlaaaagtgatgataatagttattataagttactgg 480
DB 12621 aacatatcttcttgcagaatgtlaaaagtgatgataatagttattataagttactgg 12680
QY 481 ctgtaaatgtgtctaaatataacttataatgaactaagagcttacagaacatgttgaacct 540

DB 12681 ctgtaaatgtgtctaaatataacttataatgaactaagagcttacagaacatgttgaacct 12740

QY 541 ttttactttatttgggaataaggaatgttgcacccacacatttattgctt 593
DB 12741 ttttactttatttgggaataaggaatgttgcacccacacatttattgctt 12793

RESULT 7
ID AAS29132
ID AAS29132 standard; cDNA; 1387 BP.
XX
XX AAS29132:
AC
XX 21-NOV-2001 (first entry)
DT
XX
XX cDNA encoding for human DNA-binding protein #103.
DE
XX
KW Human; DNA-binding protein; histone; chromo domain protein;
KW chromatin organisation modifier; Y-box binding protein;
KW DNA organisation; gene transcription; malignant disease;
KW autoimmune disorder; rheumatic disease; genetic abnormality;
KW infectious disease; neurological disorder; gene therapy;
KW immunomodulatory; anti-HIV; anti rheumatic; anti microbial;
KW cytoskeletal; ss.
XX
XX Homo sapiens.
OS
PN WO200155162-A1.
XX
PD 02-AUG-2001.
XX
PE 17-JAN-2001; 2001WO-0501305.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.

PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249246.
 PR 17-NOV-2000; 2000US-0249247.
 PR 17-NOV-2000; 2000US-0249249.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0256719.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PR XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM.
 XX WPI; 2001-465557/50.
 DR P-PSDB; AAU18256.
 XX
 PT Nucleic acid molecules encoding human secreted chromosomal binding
 PT proteins, used in preventing, treating or ameliorating a disorder, e.g.
 PT Alzheimer's and Parkinson's diseases and cancers -
 PT
 XX
 PS Claim 4; SEQ ID NO 113; 561pp; English.
 XX

The present invention relates to the isolation of novel DNA-binding proteins (AAU18154-AAU18281), and cDNA and genomic sequences encoding for these proteins. DNA-binding proteins such as histones, chromatin (chromatin organisation modifier) domain proteins, and Y-box binding proteins may contribute to diseases resulting from aberrant DNA organisation and/or gene transcription. The sequences of the invention are useful in screening assays to identify antagonists and/or agonists that may enhance or block activities mediated by DNA-binding proteins. Blockers of DNA-binding proteins may be useful in treating disorders such as malignant diseases (e.g. cancer), autoimmune disorders (e.g. diabetes mellitus), rheumatic diseases (e.g. rheumatoid arthritis), genetic abnormalities (e.g. cystic fibrosis), infectious diseases (e.g. HIV) and neurological disorders (e.g. Alzheimer's disease). The polynucleotide sequences of the invention may also be used in gene therapy. AAS29030-AAS29157 represent cDNA sequences encoding for novel DNA-binding proteins.
 CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1387 BP; 494 A; 165 C; 238 G; 482 T; 8 other;

Query Match 97.7%; Score 579.2; DB 22; Length 1387;
 Best Local Similarity 99.3%; Pred. No. 5.4e-111;
 Matches 569; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

QY 1 acatctatgtttacaggtcttcgtttgatagaagatagaacggaacaaactcaaaatggt 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 620 acatctatgtttacaggtcttcgtttgatagaagatagaacggaacaaactcaaaatggt 679
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 61 ggcagttcttataccagttgttagatggtttcttgcgaacatgcttgcgaagcaaacatt 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 680 ggcagttcttataccagttgttagatggtttcttgcgaacatgcttgcgaagcaaacatt 738

QY 121 tattactgttaagacacttgcttatgttgtgtgtacataatttccacaatgttata 180
|||||
Db 739 natiaactgttagaacacttgctgtatgttgtgtgtacataawttccacaatgttata 798
QY 181 attataatgtgtgtgtgaacagatgcaacttctgtgtctaaaggtgtgcagttaa 240
|||||
Db 799 attataatgtgtgtgtgaacagatgcaacttctgtgtctaaaggtgtgcagttaa 858
QY 241 aaaaaaacaacacttcttcttcaataatgcatgtgaggttttttaacttaaaac 300
|||||
Db 859 aaaaaaacaacacttcttcttcaataatgcatgtgaggttttttaacttaaaac 918
QY 301 atcaaaaattgttaaatatcattgtgtatctcagtagttaaattatcoggttatattc 360
|||||
Db 919 atcaaaaattgttaaatatcattgtgtatctcagtagttaaattatcoggttatattc 978
QY 361 cccatgaatgcatcagaactgacatttaattcaatgttctgtcgcagatgcttcttactt 420
|||||
Db 979 cccatgaatgcatcagaactgacatttaattcaatgttctgtcgcagatgcttcttactt 1038
QY 421 aacatattcttctgtcgaagtgtaaaggtatagataattagttatataaagtcttg 480
|||||
Db 1039 aacatattcttctgtcgaagtgtaaaggtatagataattagttatataaagtcttg 1098
QY 481 ctgttaattgtgtctaaatatacttatacttgcaattgaagggttaacagacatgttgaact 540
|||||
Db 1099 ctgttaattgtgtctaaatatacttatacttgcaattgaagggttaacagacatgttgaact 1158
QY 541 ttttactttatcttggaataagaatgttgcacccacacatttatgtct 593
|||||
Db 1159 ttttactttatcttggaataagaatgttgcacccacacatttatgtct 1211

RESULT 8

ABL34241 standard; DNA: 11691 BP.

AC ABL34241.

XX 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 2214.

XX Human: immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antihaemic; cytosolic; noctropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antineumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antinflamatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.

XX Homo sapiens.

OS MO200200928-A2.

PN 03-JAN-2002.

XX 02-JUL-2001; 2001MO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIDENOMICS AG.

PA Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX

PS Claim 1; SEQ ID NO 2214; 32pp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX

SQ Sequence 11691 BP; 3695 A; 49 C; 1861 G; 6086 T; 0 other;

Query Match 9.5%; Score 56.6; DB 24; Length 11691;

Best Local Similarly 46.0%; Pred. No. 0.0085;

Matches 233; Conservative 0; Mismatches 269; Indels 5; Gaps 1;

QY 79 ttgttagtattgttctcggaactgctgccaagacacatttatactgttgaacac 138
|||||
Db 5647 ttttataattctttaggaatagtgagatttttgagtatttatttagtataatttg 5706
QY 139 ttgcttaatgtgtgtgtacatatctccacaatgtatataatgatgtgtgtg 198
|||||
Db 5707 ttttttgatattttttataatattagtttagagaagatgtatttataatttataa 5766
QY 199 aacagatgcacatttctgtgtcbaaaggtgtcagttaaaaaacaacacttcc 258
|||||
Db 5767 agtatatgaatatagattattttaaaagttataaagttataataagaatgtatta 5826
QY 259 ttcaatatgcatgtgtgtgtgttttttaactttaaaacatcaaaatgttaaat 318
|||||
Db 5827 tggaaattttttatattttaaatattatataaaaaagattattttgttaatat 5886
QY 319 catgtgtatcctagtagttatataattatcgcgttatattcccatgaatgacgaac 378
|||||
Db 5887 tttttatataatataagtttgagattttgttataatatttattattgttataaa 5946
QY 379 tgacatttaattcagtgtgtgtcgcacgtcttcttactttaacaatttctttgac 438
|||||
Db 5947 gttattttaagaaggtatataattatataatattatattatagttttttttgt 6006
QY 439 aatgtaaaaggtatgaattagtttatataaggtctgcgtgaatagatgctaatt 498
|||||
Db 6007 tattt-----tttatgttaatttatttatttagatagagtttattggattagat 6061
QY 499 atacttataatgaatgaaggtctacagacatgttgaacatttttactttatggga 558
|||||
Db 6062 ttaattatataattagattttttgtgtttgtgtgttagtatttataattgtttta 6121
QY 559 ataaggaatgttgcacctccacatt 585
|||||
Db 6122 attggaagttttagtttgaatat 6148

RESULT 9

ABL33274 standard; DNA: 13038 BP.

AC ABL33274.

XX 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 1247.

XX Human: immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antihaemic; cytosolic; noctropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antineumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antinflamatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.

XX	Homo sapiens.
CS	
PN	WO200200928-A2.
XX	
PD	03-JAN-2002.
XX	
PE	02-JUL-2001; 2001WO-EP07537.
XX	
PR	30-JUN-2000; 2000DE-1032529.
PR	01-SEP-2000; 2000DE-1043826.
XX	
PA	(EPIC-) EPIGENOMICS AG.
XX	
PI	Olek A, Piepenbrock C, Berlin K;
XX	
DR	WPI; 2002-130909/17.
XX	
PT	Nucleic acid comprising fragment of chemically modified gene, useful
PT	for diagnosis and treatment of diseases associated with abnormal
XX	cytosine methylation -
XX	
PS	Claim 1; SEQ ID NO 1247; 32pp + Sequence Listing; German.
XX	
CC	The present invention provides a number of human immune system associated
CC	genes which are modified by the methylation of cytosines. The sequences
CC	can be used in the diagnosis and treatment of immune system disorders,
CC	including eye diseases such as retinopathy, neovascular glaucoma and
CC	macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC	leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC	rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC	diseases. The present sequence is a gene of the invention.
XX	
SO	Sequence 13038 BP; 4059 A; 86 C; 2457 G; 6436 T; 0 other;

	Query Match	9.3%	Score 55:	DB 24;	Length 13038;
	Best Local Similarity	52.4%;	Pred.	No. 0.016;	
Matches	144; Conservative	0;	Mismatches	130;	Indels 1; Gaps 1
Oy	77 agtctgaatgattgtttcccggaactgcctgccagacacaagaatttaactgttaga-a 135 Db	4440 atttttgttaatttgaaatggaaaaaggrrttttaaatgaaattaataatgttttagt 4499			
Oy	136 caacttcctaagtcttgytgrfacatatttccaanaatqtlatalatlaagtgttg 195 Db	4500 tgaigtgggttaaattaaattaaatgaatttatcatttaattttgtaattgatgtaattta 4559			
Oy	196 ttgaaccaagatgcacctttgttygtcetaaagvtgctgcagttaaaaaaaaaaccott 255 Db	4560 ttaaaaagfatgcttttttttttttttgttattgtgfaaataataaaaaatacatgagatt 4619			
Oy	256 tttcttcaaatagcgaatgagvgaggttttttaactttaaaaaacotcaaanatgttta 315 Db	4620 ttttttgattttctgtaagatctggttacattlatattaggagagaaatgagttgatat 4679			
Oy	316 aaatcatctgtatcatcgaqtgatttaataatcacg 350 Db	4680 aattattgittgagatatgaattattttgttcg 4714			
RESULT 10					
AC	ABL32184				
XX	ABL32184 standard; DNA; 15387 BP.				
XX	ABL32184;				
XX					
DT	26-MAR-2002 (first entry)				
DE					
XX	Human immune system associated gene SEQ ID NO: 157.				
XX					
KM	Human; Immune system disease; cytosine methylation; antisthmatic; antiartherosclerotic; antiinflammatory;cystostatic; nootropic;				

XX	neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX	anti-rheumatic; anti-arthritic; antidiabetic; antipsoritic;
KW	antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW	gene; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200200928-A2.
XX	
PD	03-JAN-2002.
XX	
PF	02-JUL-2001; 2001WO-EP07537.
XX	
PR	30-JUN-2000; 2000DE-1032529.
FR	01-SEP-2000; 2000DE-1043826.
XX	
PA	(EPIG-) EPIGENOMICS AG.
XX	
PI	Olek A, Piepenbrock C, Berlin K;
DR	WPI: 2002-130909/17.
XX	
PT	Nucleic acid comprising fragment of chemically modified gene, useful
PT	for diagnosis and treatment of diseases associated with abnormal
PT	cytosine methylation
XX	
PS	Claim 1; SEQ ID NO 157; 32pp + Sequence Listing; German.
XX	
CC	The present invention provides a number of human immune system associated
CC	genes which are modified by the methylation of cytosines. The sequences
CC	can be used in the diagnosis and treatment of immune system disorders,
CC	including eye diseases such as retinopathy, neovascular glaucoma and
CC	macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC	leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC	rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC	diseases. The present sequence is a gene of the invention.
XX	
XX	Sequence 15387 BP; 3742 A; 300 C; 3619 G; 7726 T; 0 other;

Query Match	Similarity	9.2%	Score 54.8	DB 24	Length 15387
Best Local	Similarity	45.2%	Pred. Num. 0.021		
Matches 238	Conservative	0	Mismatches 287	Indels 1	Gaps 1
QY 46	aaactcaaaaagygagcgacttctatccacagttgtgtagtatgtttctggnaaactgct	105			
Db 9796	aagaatggaattgggtgttatcttttgatttttttttttttttttttttttaagtctta	9855			
QY 106	tgccaagacacatttaactgtgttagaacactggtcttagtctgtgtgtacataatt	165			
Db 9856	ttagatataagttctgtgaacttatttttttttttttttttttttttttttttttt	9915			
QY 166	tccacaagtgttaattattatagtggtgtggaacagagatcaatctttgtgcata	225			
Db 9916	ttattagatataaagttctgttaatttttttttttttttttttttttttttttttt	9975			
QY 226	agggtgcagttaaaaaaaacaaccttcttcctcaalabgcaatgtaagtgaattt	285			
Db 9976	tttttttagatataaagtttgttaaattttttttttttttttttttttttttttttt	10035			
QY 286	tttaactttaaaaacatcaaaaattggttaaatctgtgtatcctagatgatttaatt	345			
Db 10036	atttttttattagatactaaagttgtgcaaatcttatttttttttttttttttttttt	10095			
QY 346	atcgagctatattccacatgatcatcagaactacacatttaattcatgttgcgcgc	405			
Db 10096	gttattttttttatagatvatatgaagtttggtaatttttttttttttttttttttt	10155			
QY 406	atgctcttcttact--taacatacttcctttgcagaaatgtaaaagtaatgataatagt	464			
Db 10156	attgttatatttttttttttgatattagttttgttaatttttttttttttttttttttt	10215			

OY 164 ttccacaaatgttataattatatagtgctggtgaacaggatgcaacttttgttgc 223
| | | | | ||||| | | | | |
Db 4128 tatatatataatttcataatttatatttagtaaatataaaaattacccttttattttt 4187

CC The present sequence encodes a polypeptide (Pf-SUB2) which has a
CC calcium-dependent serine-protease activity. The Pf-SUB2 gene is
CC expressed during the differentiation phase of merozoites. The protein
CC is implicated in maturation of the major surface protein 1 of
CC merozoites (MSP-142). The enzyme is also crucial for entry of the
CC parasite into erythrocytes. The polypeptides and polynucleotides are
CC used to identify inhibitors of Pf-SUB2. These inhibitors e.g. antibodies

CC are used for the detection, prevention, and treatment of malaria due to
CC Plasmodium falciparum infection.

XX Sequence 4611 BP; 1951 A; 459 C; 657 G; 1544 T; 0 other;

Query Match 9.2%; Score 54.6; DB 21; Length 4611;
Best Local Similarity 50.2%; Pred. No. 0.02;
Matches 135; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 164 ttccacaatgtataattatagtggtgtgaacagatgcaatcttctgtct 223
DB 4271 tatatatatttccattatttatttagtaataataaaccttattttt 4330
QY 224 aatgggtcgcagttaaaaaaacaaccttcttcaatgtgcatgtggt 283
DB 4331 aatatattcttgatataaatacatatatttttaatttaagtaaccttc 4390
QY 284 tttaacttaaaaacatcaaatgttaaaatcatgtgtatcctagttataa 343
DB 4391 attatatatatatatatatatatatatatatatatatttatgtatc 4450
QY 344 ttctcgctataattcccatgcatgacagactgcaatcaatcgttctcg 403
DB 4451 ttattttttaaagtagtacctattttatgtgaacaacattaccttttctgtt 4510
QY 404 ccatgctcttactttaacatttctt 432
DB 4511 ttcaatttatttatttatttattt 4539

RESULT 13

ABL33316
ID ABL33316 standard; DNA; 13574 BP.

AC ABL33316;

DT 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 1289.

KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antihaemic; cytosolic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antineumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.

XX Homo sapiens.

OS WO200200928-A2.

PN 03-JAN-2002.

PD 02-JUL-2001; 2001WO-EP07537.

PF 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIT-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
XX cytosine methylation

PS Claim 1; SEQ ID NO 1289; 32pp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.

XX Sequence 13574 BP; 3871 A; 89 C; 2670 G; 6944 T; 0 other;

Query Match 9.2%; Score 54.4; DB 24; Length 13574;
Best Local Similarity 48.0%; Pred. No. 0.025;
Matches 191; Conservative 0; Mismatches 201; Indels 6; Gaps 1;

QY 163 ttccacaatgtataattatagtggtgtgaacagatgcaatcttctgtct 222
DB 3481 tttaagaaaatgttaagatttaagtagtaagcgtaagtttctgtattgaagt 3540
QY 223 taaaggtcgcagttaaaaaaacaaccttcttcaatgtgcatgtggtgagt 282
DB 3541 gttgtttttaattaaaaataaataggtttaataatctttaaaaaatattc 3600
QY 283 tttaacttaaaaacataaatgttaaatcatgtgtatcctagttata 342
DB 3601 tgtaggtttttttaagaaaaatttataaaaaatgtaagaattttaagtaatttta 3660
QY 343 attacgctataattcccatgcatgacagactgcaatcaatcattcattgtgtc 402
DB 3661 gttagtagaataatagtagtaagtgtggaattttatttgaatatgttttg 3720
QY 403 gccatgcttcttacttaacatatttcttgc-----agatgtaaaagttaagat 456
DB 3721 gttgagtgttgtaataattatttatttatttattagatatttagatagttg 3780
QY 457 aattagttataagtgtaactgtcgttaaatgataatattacttataatgaag 516
DB 3781 tttaagttatgtgtgatagtatggtgatagaataaatttggttatataactta 3840
QY 517 ggcctacagaacatgttgaactttttactttatt 554
DB 3841 tattttagtggtgataaatttggattattttt 3878

RESULT 14

AAS45323
ID AAS45323 standard; DNA; 17848 BP.

AC AAS45323;

DT 18-DEC-2001 (first entry)

XX Chemically pretreated complementary DNA associated with cell cycle #14.

KW Cell cycle; human; Cpc dinucleotide; cytosine methylation; HIV; aging;
KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;
KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
KW immunosuppressive; antitumour; cytosolic; antiarteriosclerotic; ds;
KW PCR primer.

XX Homo sapiens.

OS WO200168911-A2.

PN 20-SEP-2001.

PD 15-MAR-2001; 2001WO-EP02945.

PF 15-MAR-2000; 2000DE-1013847.

PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.


```
Db 14971 tttttattttttt-----tttttatttgatagttttttattataagatttt 15020
QY 446 aaggtaatgataattgattataataagtgactggtgctgtaaatgataataacttc 505
Db 15021 gaattgatgaagatgagttgttctgtc--tttaattatagattagaattttttg 15078
QY 506 atgcaatttaaggcttacagaacatggttgaaacttttttactttatttggaata 561
Db 15079 attttttatcgattaaaggtaaatcaatttatttctgtatcttgtaataaa 15134
```

Search completed: May 22, 2002, 06:46:05
Job time: 8175 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 06:34:06 ; Search time 108.82 Seconds
(without alignments)
1338.547 Million cell updates/sec

Title: US-09-693-205-7_COPY_12200_12792

Perfect score: 593

Sequence: 1 acatctatgtttacagct.....accctcacattttgtctt 593

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 38353 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/prodata/1/lna/5A_COMB.seq: *
2: /cgn2_6/prodata/1/lna/5B_COMB.seq: *
3: /cgn2_6/prodata/1/lna/6A_COMB.seq: *
4: /cgn2_6/prodata/1/lna/6B_COMB.seq: *
5: /cgn2_6/prodata/1/lna/PCrUS_COMB.seq: *
6: /cgn2_6/prodata/1/lna/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47.8	8.1	5852	1	US-07-867-106-2
2	46.6	7.9	1850	3	US-08-617-860B-32
3	46.6	7.9	4098	2	US-08-605-106-4
4	44.6	7.5	19124	2	US-08-487-826B-13
5	43.2	7.3	5820	4	US-09-029-213B-7
6	41.4	7.0	665	2	US-08-883-795A-36
7	41.4	7.0	715	4	US-08-991-789A-264
8	41.4	7.0	715	4	US-09-062-451-264
9	41.4	7.0	1511	1	US-07-891-867B-8
10	41.4	7.0	1511	1	US-08-107-755A-8
11	41.4	7.0	1511	2	US-08-544-332-8
12	41.2	6.9	767	4	US-08-998-416-472
13	40.6	6.8	9048	3	US-08-973-273-4
14	39.2	6.6	7560	4	US-09-103-478-4
15	39.2	6.6	7560	4	US-09-193-931C-4
16	39.2	6.6	6152	4	US-08-973-462-1
17	38.2	6.4	1511	1	US-07-991-867B-8
18	38.2	6.4	1511	1	US-08-107-755A-8
19	38.2	6.4	1511	2	US-08-544-332-8
20	38.2	6.4	1667	1	US-08-485-284A-1
21	38.2	6.4	2781	3	US-08-749-522-4
22	38.2	6.4	6243	2	US-09-056-075-1
23	38.2	6.4	8654	1	US-08-920-812-6
24	38.2	6.4	8654	1	US-08-920-827-6
25	38.2	6.4	8654	1	US-08-921-177-6
26	38.2	6.4	8654	1	US-08-362-577C-6
27	38.2	6.4	8654	2	US-08-920-828-6

c 28	37.6	6.3	51952	3	US-08-947-823-1	Sequence 1, Appl1
c 29	37.4	6.3	660	1	US-07-991-867B-32	Sequence 32, Appl1
c 30	37.4	6.3	660	1	US-08-107-755A-32	Sequence 32, Appl1
c 31	37.4	6.3	660	2	US-08-544-332-32	Sequence 32, Appl1
c 32	37.4	6.3	4810	3	US-08-852-629-11	Sequence 11, Appl1
c 33	37.4	6.3	4838	3	US-08-852-629-15	Sequence 15, Appl1
c 34	37.4	6.3	5852	1	US-07-867-106-2	Sequence 2, Appl1
c 35	37.2	6.3	2297	2	US-09-001-826-14	Sequence 14, Appl1
c 36	37	6.2	10091	3	US-09-058-489-34	Sequence 34, Appl1
c 37	36.8	6.2	615	4	US-08-998-416-186	Sequence 186, App
c 38	36.8	6.2	731	1	US-08-451-405A-2	Sequence 2, Appl1
c 39	36.8	6.2	5253	3	US-08-714-918-19	Sequence 19, Appl1
c 40	36.8	6.2	5253	4	US-09-265-315-19	Sequence 19, Appl1
c 41	36.8	6.2	5253	4	US-09-265-315-19	Sequence 19, Appl1
c 42	36.8	6.2	5253	4	US-09-266-417-19	Sequence 19, Appl1
c 43	36.6	6.2	665	4	US-08-998-416-937	Sequence 937, App
c 44	36.6	6.2	6243	2	US-09-056-075-1	Sequence 1, Appl1
c 45	36.6	6.2	8700	2	US-08-392-625-16	Sequence 16, Appl1

ALIGNMENTS

RESULT 1
; Sequence 2, Application US/07867106
; Patent No. 5389526
; GENERAL INFORMATION:
; APPLICANT: Slade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
; TITLE OF INVENTION: Slime Moulds of the Genus Dictyostellium
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/867,106
; FILING DATE: 19920625
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PJ 7187
; APPLICATION NUMBER: PCT/AU90/00530
; FILING DATE: 02-NOV-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Feeney, Joanne Longo
; REGISTRATION NUMBER: 35,134
; REFERENCE/DOCKET NUMBER: RICE-0002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5852 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; FEATURE:
; NAME/KEY: CDS

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/605,106
FILING DATE: 23-SEPT-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP94/02935
FILING DATE: 01-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Moessner, Warren D
REGISTRATION NUMBER: 30,440
REFERENCE/DOCKET NUMBER: 235.001US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-373-6900
TELEFAX: 612-339-3061
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 4098 Base pairs
TYPE: nucleic acid
STRANDEDNESS: double stranded
TOPOLOGY: linear
MOLECULE TYPE: : DNS (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Cuphea lanceolata
IMMEDIATE SOURCE:
LIBRARY: genomic Lambda FIX II
CLONE: CITE91
FEATURE:
NAME/KEY: CDS
LOCATION: join(1797..2294, 2658..2791, 2898..3011, 3132
LOCATION: ..3303, 3391..3459, 3672..3941)
FEATURE:
NAME/KEY: Startcodon
LOCATION: 1797..1799
FEATURE:
NAME/KEY: exon II
LOCATION: 1787..2294
FEATURE:
NAME/KEY: Intron II
LOCATION: 2295..2657
FEATURE:
NAME/KEY: exon III
LOCATION: 2658..2791
FEATURE:
NAME/KEY: Intron III
LOCATION: 2792..2897
FEATURE:
NAME/KEY: exon IV
LOCATION: 2898..3011
FEATURE:
NAME/KEY: Intron IV
LOCATION: 3012..3131
FEATURE:
NAME/KEY: exon V
LOCATION: 3132..3303
FEATURE:
NAME/KEY: Intron V
LOCATION: 3304..3390
FEATURE:
NAME/KEY: exon VI
LOCATION: 3391..3459
FEATURE:
NAME/KEY: Intron VI
LOCATION: 3460..3671
FEATURE:
NAME/KEY: exon VII
LOCATION: 3672..3941

FEATURE:
NAME/KEY: Stopcodon
LOCATION: 3942..3944
US-08-605-106-4

Query Match 7.9%; Score 46.6; DB 2; Length 4098;
Best Local Similarity 44.8%; Pred. No. 0.019;
Matches 178; Conservative 0; Mismatches 219; Indels 0; Gaps 0;

QY 153 tgtgtacatattccacaagtgtataatcatalatggtgtgagcagatgcaatc 212
D 182 TATTGAATATTTTGAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 241
QY 213 ttgtgtgtcctaaggtgtcagttcaaaaaaacacaccttcttcaatgcat 272
D 242 TTTTAAAAAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 301
QY 273 gtagtgaagtttttaactaaacacaaacaaatgtaaaatcattgtatccta 332
D 302 AAAATTAGTTTATTTTATTTTAAATTTTGAATTTTAAATTTTATTTTGGTTT 361
QY 333 gtagttataattcgcgctataatcccatgaatgacagactgacattaatca 392
D 362 AAAATATATTTTAAAGTTTAAATTTTAAATTTTGAATTTTGAATTTTGAATTTT 421
QY 393 tgttgcgcgcagctctcttacttaacatatcttcttcagaagtgaagaagtaa 452
D 422 TTGGAGATTAACCGAGATTTATATATATATATATATATATATATATATATTC 481
QY 453 tgataattagttatataagtgctacgtcgtgtaaatgtaaatcattatgcaat 512
D 482 GTCCATTTGGTTAAACCAAGTAGTTTGTAAACAGATGATGATGATGATGATGAT 541
QY 513 taagggtctacagaacatgttgaacctttttactt 549
D 542 GAAAGTTTAAAGCAACGAAGCATATTTTATTTT 578

RESULT 4
US-08-487-826B-13/C
Sequence 13, Application US/08487826B
Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Sim, Kim I.
APPLICANT: Chltnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Krobbe Mariens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelien, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI

STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/062,451
FILING DATE: 04-APR-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121,419C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 264:
SEQUENCE CHARACTERISTICS:
LENGTH: 715 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-062-451-264

Query Match 7.0%; Score 41.4; DB 4; Length 715;
Best Local Similarity 48.7%; Pred. No. 0.22; Mismatches 117; Indels 0; Gaps 0;

Matches 111; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 326 tctacagtagtataatcgcgttattccccaatgaatgacgaagtcatt 385
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 157 tttattttaaataatgatttttttgagattcattttatagattttatcacta 216
QY 386 taattcattgttctgcgcacatgcttcttacttaacatattcttgcagaatgtaa 445
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 217 ttgtatttctttgtttgttggtcttctgcatcttcaatgcatcttaaacatcaca 276
QY 446 aagtgatgataatgattataatgtagtgcgtgtaaaatgacgaataatactt 505
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 277 atctattttcaaatatattatcatttaacattatgtaataatgtaaacacacattata 336
QY 506 atgcgaatgaagcttaccagaacatgtgaacttttttactttat 553
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 337 ttcccatttctcccttccatccatgintgcatatttttttctttat 384

RESULT 9
US-07-991-867B-8
Sequence 8, Application US/07991867B
Patent No. 5476781
GENERAL INFORMATION:
APPLICANT: Moyer, Richard W.
APPLICANT: Hall, Richard L.
APPLICANT: Gruidl, Michael E.
TITLE OF INVENTION: No. 5476781el Entomopoxvirus Expression System
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanichk
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/991,867B

FILING DATE: 12-DEC-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 92/14818
FILING DATE: 12-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/827,685
FILING DATE: 30-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/657,584
FILING DATE: 19-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanichk, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: UF114.C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1511 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Amsacta moorei entemopoxvirus
FEATURE:
NAME/KEY: CDS
LOCATION: complement (18..218)
FEATURE:
NAME/KEY: CDS
LOCATION: complement (234..782)
FEATURE:
NAME/KEY: CDS
LOCATION: 852..1511
US-07-991-867B-8

Query Match 7.0%; Score 41.4; DB 1; Length 1511;
Best Local Similarity 47.2%; Pred. No. 0.25; Mismatches 176; Indels 2; Gaps 1;

Matches 159; Conservative 0; Mismatches 176; Indels 2; Gaps 1;

QY 176 ttaattatataatgtagtggtagaagaatcctttgtgtcctaagggtcgta 235
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 553 tttattcagcattatgatttcatatttaattttttgtttatgtaataatttttta 612
QY 236 gttaaaaaaaacaaacctttcttcaataltggaatgtagtggagtttttaactta 295
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 613 ttttaattattttccgcacatgattttattttttttattataaactatttatattata 672
QY 296 aaaaactcaaaaatltgttaaaatcattggtctcagtagtattataatcaggttat 355
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 673 tga--gttatttaattacacattttttgattgatttaataattatttttttgcgacata 730
QY 356 attccccaatgaatgtagcgaacatgacatttaataatcattgttctgcacatgctctt 415
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 731 attctctgtttgttgcgaagaaacatnagacacattttttattttatcgcacatttttttt 790
QY 416 acttaacatattcttcttgcagaatgtaaaagtaagtaataatgattatataagtg 475
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 791 attatttggattatttttttttcaaaaaaaatttaatacagaataaaaaaaattttatca 850
QY 476 actggtgttaaatgtagtcaaatatactttatgcaat 512
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 851 aatgattttacttaatttcgattatttttttaatttaatt 887

RESULT 10
US-08-107-755A-8
Sequence 8, Application US/08107755A
Patent No. 5721352
GENERAL INFORMATION:

```

? APPLICANT: Moyer, Richard W.
? APPLICANT: Hall, Richard L.
? APPLICANT: Gruidl, Michael E.
? TITLE OF INVENTION: No. 5721352el Entomopoxvirus Expression System
? NUMBER OF SEQUENCES: 40
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: David R. Saliwanchik
? STREET: 2421 N.W. 41st Street, Suite A-1
? CITY: Gainesville
? STATE: Florida
? COUNTRY: U.S.A.
? ZIP: 32606
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/107,755A
? FILING DATE: 19-AUG-1993
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/827,658
? FILING DATE: 30-JAN-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/657,584
? FILING DATE: 19-FEB-1991
? ATTORNEY/AGENT INFORMATION:
? NAME: Saliwanchik, David R.
? REGISTRATION NUMBER: 31,794
? REFERENCE/DOCKET NUMBER: UF114.C2
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (904) 375-8100
? TELEFAX: (904) 372-5800
? INFORMATION FOR SEQ ID NO: 8:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1511 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: unknown
? MOLECULE TYPE: DNA (genomic)
? ORIGINAL SOURCE:
? ORGANISM: Amsacta moorei entemopoxvirus
? FEATURE:
? NAME/KEY: CDS
? LOCATION: complement (18..218)
? FEATURE:
? NAME/KEY: CDS
? LOCATION: complement (234..782)
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 852..1511
? US-08-107-755A-8

Query Match 7 0%; Score 41.4; DB 1; Length 1511;
Best Local Similarity 47.2%; Pred. No. 0.25;
Matches 159; Conservative 0; Mismatches 176; Indels 2; Gaps 1;

QY 176 ttataattatagtggtgacagagatgcacatcttctgtgtctaaagtcgcga 235
DB 553 TTATCTCGATATTGATTCATTAATAATTATTTTAAATGTAATAATATCTTTA 612
QY 236 gtaaaaaaaacacaccttcttcaataatgcatgtagtgagtttttaactta 295
DB 613 TTTAAATATTTCCGTCATGATTTATATATTTTATTTAAATCTATATCTATATTA 672
QY 296 aaaaatcaaaatgttaaaatcatgtgttatctagttatataataatgcgctat 355
DB 673 TGA-GTTAATATTACATATTTGATTAGATAAAATATCTAATATTTTCGATCA 730
QY 356 attcccaagaatgacgaactgaacttaactatgcttgcgcacatgctctt 415

```

```

DB 731 ATTCTGTTGTTTCCAGAAACATAGACCATTATTAAATCTATGACATTTTTTTT 790
QY 416 acttaacataatctcttcgagaatgtaaaaggaatgataatgatatagtc 475
DB 791 ATTATTTGATATATTTTTCATAAAAAAATTAATCAATGAAAAAATAAATATCA 850
QY 476 actgagctaaatgactaaataatcttatgcaat 512
DB 851 AATGATTTACTAATCTCATATATTTTAATAAT 887

RESULT 11
US-08-544-332-8
? Sequence 8, Application US/08544332
? Patent No. 5935777
? GENERAL INFORMATION:
? APPLICANT: Moyer, Richard W.
? APPLICANT: Hall, Richard L.
? APPLICANT: Gruidl, Michael E.
? TITLE OF INVENTION: No. 5935777el Entomopoxvirus Expression System
? NUMBER OF SEQUENCES: 77
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Gerard H. Bencen
? STREET: 2421 N.W. 41st Street, Suite A-1
? CITY: Gainesville
? STATE: FL
? COUNTRY: USA
? ZIP: 32606
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: IBM PC compatible
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/544,332
? FILING DATE:
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/991,867
? FILING DATE: 07-DEC-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/107,755
? FILING DATE: 19-AUG-1993
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: WO 92/14818
? FILING DATE: 12-FEB-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/827,685
? FILING DATE: 30-JAN-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/657,584
? FILING DATE: 19-FEB-1991
? ATTORNEY/AGENT INFORMATION:
? NAME: Bencen, Gerard H.
? REGISTRATION NUMBER: 35,746
? REFERENCE/DOCKET NUMBER: UF114.C4
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 904-375-8100
? TELEFAX: 904-372-5800
? INFORMATION FOR SEQ ID NO: 8:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1511 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: unknown
? MOLECULE TYPE: DNA (genomic)
? ORIGINAL SOURCE:
? ORGANISM: Amsacta moorei entemopoxvirus
? FEATURE:
? NAME/KEY: CDS
? LOCATION: complement (18..218)
? FEATURE:
? NAME/KEY: CDS

```

LOCATION: complement (234..782)
 FEATURE: CDS
 NAME/KEY: 852..1511
 LOCATION: 852..1511
 US-08-544-332-8

Query Match 7.0%; Score 41.4; DB 2; Length 1511;
 Best Local Similarity 47.2%; Pred. No. 0.25;
 Matches 159; Conservative 0; Mismatches 176; Indels 2; Gaps 1;

QY 176 ttataattatagtggtgagacagatgcaatcttcttgctcgaaggctgca 235
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 553 tttatcagcatattgattcatttaattatttcttttaattatttctttaa 612
 QY 236 gttaaaaaaaacaaccccttcttccatattgcatgtagtggagcttttttaactta 295
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 613 ttttaattattttccgcatgattttattatttttttataaattctattatttata 672
 QY 296 aaacatcaaaaattgtaaatcatctgcttactagtagttaaatcatcgcttat 355
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 673 tga--gtttattattacacattttttgatttagatttaattatttctatttttccatca 730
 QY 356 attccccatgaatgacgaactgacatttaattcaatgctgctcgcacatgcttctt 415
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 731 attctgtgttttccagaaaacatagacacattttatttctattcagacatttttttt 790
 QY 416 actttaacattctcttcttcagagaatgtaaaagtgatataattgcttataatgctc 475
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 791 attatttgaatttttttttcaaaaaaatttaavcaatgaaaaaatttaattatca 850
 QY 476 actgctgtaaatgacgaatataacttaacttaact 512
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 851 aatgattttacttaattctgattatttttaatttaatt 887

RESULT 12
 US-08-998-416-472

; Sequence 472, Application US/08998416
 ; Patent No. 6239264

GENERAL INFORMATION:
 APPLICANT: Philippsen, Peter
 APPLICANT: Pohlmann, Rainer
 APPLICANT: Steiner, Sabine
 APPLICANT: Mohr, Christine
 APPLICANT: Wendland, Jürgen
 APPLICANT: Knechtle, Philipp
 APPLICANT: Redischung, Corinne
 TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPPII
 NUMBER OF SEQUENCES: 1152
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 6239264artis Corporation
 STREET: 3054 Cornwallis Road
 CITY: Research Triangle Park
 STATE: No. 6239264th Carolina
 COUNTRY: USA
 ZIP: 27709

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/998,416
 FILING DATE: 24-DEC-1997

CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: CH 0016/97
 FILING DATE: 31-DEC-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Meigs, J. Timothy
 REGISTRATION NUMBER: 38,241

REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8587
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 472:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 767 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORGANISM SOURCE:
 ORGANISM: PAG13378P
 US-08-998-416-472

Query Match 6.9%; Score 41.2; DB 4; Length 767;
 Best Local Similarity 45.1%; Pred. No. 0.25;
 Matches 189; Conservative 0; Mismatches 220; Indels 10; Gaps 1;

QY 67 tcttattaccagtgtagtattgttctcggaaactgcttgcacaagaacatttaaa 126
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 291 ttttaattatttaattatccattttttatttaatttaattgattatttattttattttat 350
 QY 127 ctgtagaacactgcttattatgttggtgacataattccacaagtltataattat 186
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 351 gtgatattattattttattattattattttatttttacttctgcatattattattttat 410
 QY 187 atagtggttggaacagatgcaatcttcttgcttcaaaagtgctgcagttaaaaaa 246
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 411 taaatgtaccttcattatatt 470
 QY 247 aacaaccttcttcttcaatagcagtgtagtgagttttttaaacttaaaacatcaaa 306
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 471 tccctttatttgatatttacttacttaatttttttacttaattattttattttatgaatt 530
 QY 307 aattgttaaatcatctgltatclagtagttat-----aattacggccttata 356
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 531 acttaaatcattatattttatt 590
 QY 357 tttccccatgaatgacgaactgacatttaattcaatgcttctgcacatgcttctta 416
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 591 ttttttaattatt 650
 QY 417 cttaacatacttcttcttcagaaatgtaaaagtgatagttatattataatgctc 475
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 651 gatnattattatttcttt 709

RESULT 13

US-08-973-273-4/C
 ; Sequence 4, Application US/08973273
 ; Patent No. 6140085

GENERAL INFORMATION:
 APPLICANT: Dean, Caroline
 APPLICANT: MacKnight, Richard C
 APPLICANT: Bancroft, Ian
 APPLICANT: Lister, Clare K
 TITLE OF INVENTION: Genetic Control of Flowering
 NUMBER OF SEQUENCES: 31
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Nixon & Vanderhye P.C.
 STREET: 1100 No. 6140085th Glebe Road, 8th floor
 CITY: Arlington

STATE: Virginia
 COUNTRY: United States of America
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/973,273

FILING DATE: 01-DEC-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/GB96/01332
 FILING DATE: 03-JUN-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB 9511196.9
 FILING DATE: 02-JUN-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Ms Mary J Wilson
 REGISTRATION NUMBER: 32,955
 REFERENCE/DOCKET NUMBER: 620-29
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 816-4000
 TELEFAX: (703) 816-4100
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 9048 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Brassica
 US-08-973-273-4

Query Match 6.8%; Score 40.6; DB 3; Length 9048;
 Best Local Similarity 43.3%; Pred. No. 0.55;
 Matches 190; Conservative 0; Mismatches 249; Indels 0; Gaps 0;

QY 116 acatttattactgtagaacactgcttattggtggtacataattccacaatg 175
 DB 818 ACTTATGTTGATTAAAAAACTCTTATTTTGTATTAAGATCAATCATAT 759
 QY 176 ttataattatagtggtgtagaacagatgcacattcttgtgtcctaagtgctgca 235
 DB 758 TTATATATCAATTAATTTAGTAAATTTGCTTATTTGTCATGTTTATTAAGGTTTTAGC 699
 QY 236 gttaaaaaaaacaccttcttccaatagcatgtagtgaggtttttaactta 295
 DB 698 TAATCATGATTATTATTATTATTAGCTAGCTCCTATTTATTTAAAAATA 639
 QY 296 aaaaactcaaaaatcgtlaaatcatgltacgttagtattataatcggcttat 355
 DB 638 CCTTAAGATCTTATATATATTAAAAATGCAATTTTATTAGTAATTAATAATTTATAT 579
 QY 356 attcccatgaaatgcagaactgacattatcatgltgtctcgccatgctcttc 415
 DB 578 TTATATTTAAATTTAGTAAATTTCTTATTTGTCATATTATTAGGTTTAAAGACTTT 519
 QY 416 acttaacatattctcttcgagaatgtaaaggtaagtaattagttatataagtg 475
 DB 518 TACTAGTATTGATTATTATTATTATTCTAGCTGATGTTCCGTAATGCTTATTTAAAA 459
 QY 476 actgagctglaaatgagctaaatatactatgcgaatgaaggcttcagaacatg 535
 DB 458 CAATACCTTAATGAATTTATTAATTAAGAAAGCAATTTTATTAATCACTGAATAATAT 399
 QY 536 aacttttttaactttatt 554
 DB 398 ATGTTATATTAAATTAAT 380

RESULT 14
 US-09-103-478-4/C
 Sequence 4, Application US/09103478
 Patent No. 6235975
 GENERAL INFORMATION:
 APPLICANT: Harada, John
 APPLICANT: Lotan, Tamar

APPLICANT: Ohto, Masa-aki
 APPLICANT: Goldberg, Robert B.
 APPLICANT: Fischer, Robert L.
 TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
 NUMBER OF SEQUENCES: 29
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/103,478
 FILING DATE: 24-JUN-1998
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 09/026,221
 FILING DATE: 19-FEB-1998
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/804,534
 FILING DATE: 21-FEB-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Elmhorn, Gregory P.
 REGISTRATION NUMBER: 38,440
 REFERENCE/DOCKET NUMBER: 023070-07611US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 7560 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-09-103-478-4

Query Match 6.6%; Score 39.2; DB 4; Length 7560;
 Best Local Similarity 52.7%; Pred. No. 1.1;
 Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 271 atgtagtggagtttttaactttaaaaacatacaaatggttaaaatcttggtatc 330
 DB 665 ATTTTGTAAATTTTACGTTTACTAACATCATATTTTTCGATTTAGNGATATGNT 606
 QY 331 tagtagttaaatatcgccttatattcccatgaaatgacgaactgacatttaatt 390
 DB 605 GNAAAATTTTCAATCGGNANGATATTAATTAATGCGCTCAATGAAAAAACAATATTCCT 546
 QY 391 catgttctgcgcacatgctctta 416
 DB 545 GAGGTAAATTTTGAGATATCTTTA 520

RESULT 15
 US-09-193-931C-4/C
 Sequence 4, Application US/09193931C
 Patent No. 6320102
 GENERAL INFORMATION:
 APPLICANT: Harada, John
 APPLICANT: Lotan, Tamar
 APPLICANT: Ohto, Masa-aki
 APPLICANT: Goldberg, Robert B.
 APPLICANT: Fischer, Robert L.
 APPLICANT: The Regents of the University of California
 TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses

```

; FILE REFERENCE: 023070-077620
; CURRENT APPLICATION NUMBER: US/09/193,931C
; CURRENT FILING DATE: 1998-11-17
; PRIOR APPLICATION NUMBER: US 09/103,478
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: US 09/026,221
; PRIOR FILING DATE: 1998-02-19
; PRIOR APPLICATION NUMBER: US 08/804,534
; PRIOR FILING DATE: 1997-02-21
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 4
; LENGTH: 7560
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: 7.4 kb genomic wild-type fragment containing LEC1
; NAME/KEY: misc_feature
; LOCATION: (2430)..(5824)
; OTHER INFORMATION: corresponds to SEQ ID NO:3
; NAME/KEY: promoter
; LOCATION: (2430)..(4427)
; OTHER INFORMATION: corresponds to LEC1 promoter in SEQ ID NO:3
; NAME/KEY: CDS
; LOCATION: (4427)..(5054)
; OTHER INFORMATION: LEAFY COTYLEDON1 (LEC1)
; NAME/KEY: modified base
; LOCATION: (1)..(7560)
; OTHER INFORMATION: n = g, a, c or t
; US-09-193-931C-4

```

```

Query Match          6.6%; Score 39.2; DB 4; Length 7560;
Best Local Similarity 52.7%; Pred.No.1.1;
Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 271 atgtagtggaagtttttaacttaaaacatcaaaaatgtttaaatacatgtgtatc 330
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 665 ATTTGTGTAATTTTTCGTTTACTACATCATATTTTGTGATATTAGNGATATGNT 606

QY 331 tagtagttcataatcgcgcttataatccccaatgaatgatacagaactgacatttaatt 390
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 605 GNAATAATTTTCNATCGNANNGTATATATGCGCTCAATTGAAAAAACAATAATTTCC 546

QY 391 catgttcgcgcacatgccttcta 416
   ||||| ||||| ||||| |||||
DB 545 GAGGTTAATTTTGTAGATATCTTTA 520

```

Search completed: May 22, 2002, 06:34:29
Job time: 7574 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 06:46:05 ; Search time 463.88 Seconds
(without alignments)
74.024 Million cell updates/sec

Title: US-09-693-205-47

Perfect score: 20
Sequence: 1 gtgaatggccacttgcact 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_032802:*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	20	AAH20154	Human spastin ORF
2	20	100.0	12792	AAH20176	Human spastin ORF
3	20	100.0	12793	AAH20174	Human spastin ORF
4	20	100.0	12793	AAH20178	Human spastin ORF
5	20	100.0	12793	AAH20179	Human spastin ORF
6	20	100.0	12793	AAH20182	Human spastin ORF
7	18.4	92.0	11453	AAH20175	Human spastin ORF
8	17.4	87.0	901	ABA08683	Human PRO257 homol
9	16.4	82.0	1986	AAH83995	DNA encoding novel

C 10	16.4	82.0	2206	23	AAH79251	DNA encoding novel
C 11	16.4	82.0	2206	23	AAH82805	DNA encoding novel
C 12	16.4	82.0	2206	23	AAH83970	DNA encoding novel
C 13	16.4	82.0	2206	23	AAH85113	DNA encoding novel
C 14	16.4	82.0	2206	23	AAH88179	DNA encoding novel
C 15	16.4	82.0	2206	23	AAH89592	DNA encoding novel
C 16	16.4	82.0	4254	23	AAH83921	DNA encoding novel
C 17	16.4	82.0	4254	23	AAH76065	DNA encoding novel
C 18	16.4	82.0	4345	23	AAH89596	DNA encoding novel
C 19	16.4	82.0	6458	23	AAH84050	DNA encoding novel
C 20	15.8	79.0	197	21	AAH20851	Human breast cancer
C 21	15.8	79.0	336	22	AAH11827	Human breast cancer
C 22	15.8	79.0	337	22	AAH20717	Nicotiana tabacum
C 23	15.8	79.0	927	24	AAH01664	Nicotiana tabacum
C 24	15.8	79.0	1210	24	AAH01665	DNA encoding novel
C 25	15.8	79.0	10175	23	AAH70589	Human reproductive
C 26	15.8	79.0	12970	22	AAH05001	Human cDNA clone (
C 27	15.4	77.0	590	22	AAH09497	Human cDNA clone (
C 28	15.4	77.0	1449	22	AAH15402	Hydroxylamine oxid
C 29	15.4	77.0	5105	15	AAH78177	Drosophila melanog
C 30	15.4	77.0	8095	23	AAH24026	Drosophila melanog
C 31	15.4	77.0	9182	23	AAH04532	Drosophila melanog
C 32	15.4	77.0	26767	23	AAH08654	Drosophila melanog
C 33	15.4	77.0	26960	23	AAH08712	Drosophila melanog
C 34	15.2	76.0	413	22	AAH89167	Human polynucleoti
C 35	15.2	76.0	532	22	AAH11396	Human cDNA clone (
C 36	15.2	76.0	710	24	AAH61928	Porcine muscular s
C 37	15.2	76.0	768	17	AAH16240	Porcine reproductive
C 38	15.2	76.0	771	22	AAH26334	Flea saliva proteol
C 39	15.2	76.0	1068	19	AAH73434	Flea saliva proteol
C 40	15.2	76.0	1068	19	AAH73435	Flea saliva proteol
C 41	15.2	76.0	1071	19	AAH73436	Flea saliva proteol
C 42	15.2	76.0	1125	19	AAH73432	Flea saliva proteol
C 43	15.2	76.0	1125	19	AAH73433	Flea saliva proteol
C 44	15.2	76.0	1300	19	AAH73430	Flea saliva proteol
C 45	15.2	76.0	1300	19	AAH73431	Flea saliva proteol

ALIGNMENTS

RESULT 1

ID AAH20154 standard; DNA: 20 BP.

AAH20154:

09-AUG-2001 (first entry)

Human spastin ORF PCR primer SEQ ID NO:47.

Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
atrophy of upper cerebellar vermis; absence of Purkinje cell;
abnormal neuronal lipid storage; genetic disorder; characterisation;
PCR primer: ss.

Human sapiens.
OS Synthetic.
XX

WO200129266-A2.

26-APR-2001.

20-OCT-2000; 2000WO-US29130.

20-OCT-1999; 99US-0160588.

(UYMC-) UNIV MCGILL.

(HOPI-) HOPITAL SAINTE-JUSTINE.

PI Hudson TJ, Engert J, Richter A;
 XX WPI; 2001-308494/32.
 XX
 DR New isolated polynucleotide, encoding spastin gene, and polypeptides,
 XX useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -
 XX
 PS Claim 23; Fig 7; 76pp; English.
 XX
 CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with Spastin
 CC protein or spastin gene expression or activity. AAH20122 to AAH20173
 CC represent specifically claimed primers which can be used in diagnostic
 CC methods from the present invention.
 XX
 SQ Sequence 20 BP; 4 A; 5 C; 5 G; 6 T; 0 other;
 XX
 Query Match 100.0%; Score 20; DB 22; Length 20;
 Best Local Similarity 100.0%; Pred. No. 0.3; Mismatches 0; Gaps 0;
 Matches 20; Conservative 0; Indels 0; Gaps 0;
 QY 1 gtgaatggccacttgctcact 20
 ||||||||||||||||||
 Db 1 gtgaatggccacttgctcact 20
 XX
 RESULT 2
 AAH20176
 ID AAH20176 standard; DNA; 12792 BP.
 XX
 AC AAH20176;
 XX
 DT 09-AUG-2001 (first entry)
 XX
 DE Human mutated spastin nucleotide sequence SEQ ID NO:7.
 XX
 KW Human: mouse; spastin; ARSACS; chromosome 13q11; identification;
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FH 77..6604
 FT CDS /*tag= a
 FT /product= "mutated spastin"
 XX

PN W0200129266-A2.
 XX
 XX 26-APR-2001.
 XX
 PF 20-OCT-2000; 2000MO-US29130.
 XX
 PR 20-OCT-1999; 9905-0160588.
 XX
 PA (UYMC-) UNIV MCGILL.
 PA (HOPIT-) HOPITAL SAINTE-JUSTINE.
 XX
 PI Hudson TJ, Engert J, Richter A;
 XX WPI; 2001-308494/32.
 DR P-PSDB; AAB97821.
 XX
 PS Claim 1; Page -: 76pp; English.
 XX
 CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with Spastin
 CC protein or spastin gene expression or activity. The present sequence
 CC encodes a mutated human spastin.
 CC N.B. The present sequence is not given in the present specification but
 CC is derived from the human spastin nucleotide sequence (AAH20174) as
 CC stated on page 14.
 XX
 SQ Sequence 12792 BP; 4163 A; 2256 C; 2487 G; 3886 T; 0 other;
 XX
 Query Match 100.0%; Score 20; DB 22; Length 12792;
 Best Local Similarity 100.0%; Pred. No. 0.95; Mismatches 0; Gaps 0;
 Matches 20; Conservative 0; Indels 0; Gaps 0;
 QY 1 gtgaatggccacttgctcact 20
 ||||||||||||||||||
 Db 6473 gtgaatggccacttgctcact 6492
 XX
 RESULT 3
 AAH20174
 ID AAH20174 standard; DNA; 12793 BP.
 XX
 AC AAH20174;
 XX
 DT 09-AUG-2001 (first entry)
 XX
 DE Human spastin nucleotide sequence SEQ ID NO:1.
 XX
 KW Human: mouse; spastin; ARSACS; chromosome 13q11; identification;
 KW

KM autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KM neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KM reduced motor nerve velocity; hypermyelination of retinal nerve fibres;
 KM atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KM abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
 OS Homo sapiens.

PH Key Location/Qualifiers
 FT CDS 77..11566
 FT /tag= a
 FT /product= "spastin"

PN W0200129266-A2.

PD 26-APR-2001.

PF 20-OCT-2000; 2000MO-US29130.

PR 20-OCT-1999; 99US-0160588.

PA (UYMC-) UNIV MCGILL.
 PA (HOP1-) HOPITAL SAINTE-JUSTINE.

PI Hudson TJ, Engert J, Richter A;

DR WPI: 2001-308494/32.
 DR P-PSDB; AAB97819.

XX New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -

PS Claim 1: Fig 9; 76pp; English.

XX The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSA) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with spastin
 CC protein or spastin gene expression or activity. The present sequence
 CC encodes human spastin as given in the present invention.

XX Sequence 12793 BP; 4163 A; 2256 C; 2487 G; 3887 T; 0 other;

Query Match 100.0%; Score 20; DB 22; Length 12793;
 Best Local Similarity 100.0%; Pred. No. 0.95;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgaatgcacatttcgact 20
 ||||||||||||||||

DB 6473 gtgaatgcacatttcgact 6492

RESULT 4

AAH20178
 ID AAH20178 standard; DNA: 12793 BP.

XX AAH20178;

DT 09-AUG-2001 (first entry)

XX Human mutated spastin nucleotide sequence SEQ ID NO:11.

XX Human: mouse; spastin; ARSA; chromosome 13q11; identification;
 KM autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KM neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KM reduced motor nerve velocity; hypermyelination of retinal nerve fibres;
 KM atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KM abnormal neuronal lipid storage; genetic disorder; characterisation; ds.

XX Homo sapiens.

OS Synthetic.

PN W0200129266-A2.

PD 26-APR-2001.

PF 20-OCT-2000; 2000MO-US29130.

PR 20-OCT-1999; 99US-0160588.

PA (UYMC-) UNIV MCGILL.
 PA (HOP1-) HOPITAL SAINTE-JUSTINE.

PI Hudson TJ, Engert J, Richter A;

DR WPI: 2001-308494/32.

XX New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -

PS Claim 1: Page -; 76pp; English.

XX The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSA) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with spastin
 CC protein or spastin gene expression or activity. The present sequence
 CC represents a mutated human spastin gene from the present invention.
 CC N.B. The present sequence is not given in the present specification but
 CC is derived from the human spastin nucleotide sequence (AAH20174) as
 CC stated on page 14.

XX Sequence 12793 BP; 4163 A; 2255 C; 2487 G; 3888 T; 0 other;

Query Match 100.0%; Score 20; DB 22; Length 12793;
 Best Local Similarity 100.0%; Pred. No. 0.95;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 gtgaatgcccattgcact 20
 ||||||||||||||||
 Db 6473 gtgaatgcccattgcact 6492

RESULT 5
 AAH20179
 ID AAH20179 standard; DNA; 12793 BP.

AC AAH20179;

DT 09-AUG-2001 (first entry)

DE Human mutated spastin nucleotide sequence SEQ ID NO:12.

KM Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
 KM autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KM neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KM reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KM atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KM abnormal neuronal lipid storage; genetic disorder; characterisation; ds.

OS Homo sapiens.
 OS Synthetic.

PN WO200129266-A2.

PD 26-APR-2001.

PF 20-OCT-2000; 2000WO-US29130.

PR 20-OCT-1999; 99US-0160588.

PA (UYMC-) UNIV MCGILL.
 PA (HOP-) HOPITAL SAINTE-JUSTINE.

PI Hudson TJ, Engert J, Richter A;

DR WPI; 2001-308494/32.

PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -

PS Claim 1: Page -: 76pp; English.

XX The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with spastin
 CC protein or spastin gene expression or activity. The present sequence
 CC encodes a mutated human spastin.
 CC N.B. The present sequence is not given in the present specification but

CC is derived from the human spastin nucleotide sequence (AAH20174) as
 CC stated on page 14.
 XX
 SQ Sequence 12793 BP; 4163 A; 2257 C; 2487 G; 3886 T; 0 other;

Query Match 100.0%; Score 20; DB 22; Length 12793;
 Best Local Similarity 100.0%; Pred. No. 0.95;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgaatgcccattgcact 20
 ||||||||||||||||
 Db 6473 gtgaatgcccattgcact 6492

RESULT 6

ID AAH20182 standard; DNA; 12793 BP.

AC AAH20182;

DT 09-AUG-2001 (first entry)

DE Human mutated spastin nucleotide sequence SEQ ID NO:15.

XX Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
 KM autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KM neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KM reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KM atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KM abnormal neuronal lipid storage; genetic disorder; characterisation; ds.

OS Homo sapiens.
 OS Synthetic.

EH Key Location/Qualifiers
 FT 77..11566
 FT /*lag- a
 FT /*product= "mutated spastin"

PN WO200129266-A2.

PD 26-APR-2001.

PF 20-OCT-2000; 2000WO-US29130.

PR 20-OCT-1999; 99US-0160588.

PA (UYMC-) UNIV MCGILL.
 PA (HOP-) HOPITAL SAINTE-JUSTINE.

PI Hudson TJ, Engert J, Richter A;

DR WPI; 2001-308494/32.

DR P-PDB: AAB97823.

PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -

PS Claim 1: Page -: 76pp; English.

XX The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,

CC atrophy of upper cerebellar vermis; absence of Purkinje cells and
CC abnormal neuronal lipid storage. (I) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (I). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with Spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes a mutated human spastin.
CC N.B. The present sequence is not given in the present specification but
CC is derived from the human spastin nucleotide sequence (AAH20174) as
CC stated on page 14.
XX
SQ Sequence 12793 BP; 4163 A; 2257 C; 2487 G; 3886 T; 0 other;

Query Match 100.0%; Score 20; DB 22; Length 12793;
Best Local Similarity 100.0%; Pred. No. 0.95;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtgaatggcaccattgcact 20
|||||
Db 6473 gtgaatggcaccattgcact 6492

RESULT 7
AAH20175
ID AAH20175 standard; DNA: 11493 BP.
XX
AC AAH20175;
XX
DT 09-AUG-2001 (first entry)
XX
DE Mouse spastin nucleotide sequence SEQ ID NO:3.

XX Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 1..11493
FT /*tag= a
FT /product= "spastin"
XX
PN MO200129266-A2.
XX
PD 26-APR-2001.
XX
PF 20-OCT-2000; 2000WO-US29130.
XX
PR 20-OCT-1999; 99US-0160588.
XX
PA (UYMC-) UNIV MCGILL.
PA (HOPIT-) HOPITAL SAINTE-JUSTINE.
XX
PI Hudson TJ, Engert J, Richter A;
XX
DR WPL: 2001-308494/32.
DR P-PSDB; AAB97820.
XX
PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -

XX
PS Claim 1; Fig 8; 76pp; English.

XX The present invention describes human and mouse spastin, and mutated
XX human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
XX (ARSACS)) gene sequences (I). The spastin gene has been mapped to
XX chromosome 13q11. (I) have neuroprotective activities and can be used in
XX gene therapy and as a spastin polypeptide agonists. (I), their fragments
XX or their complements can be useful for assaying the presence of a nucleic
XX acid molecule in a sample. (I) is useful for diagnosing or aiding in the
XX diagnosis of an early onset neurodegenerative disease in an individual.
XX The neurodegenerative disease comprises reduced sensory nerve conduction,
XX reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
XX atrophy of upper cerebellar vermis, absence of Purkinje cells and
XX abnormal neuronal lipid storage. (I) can also be used to produce
XX antisense nucleic acids, is useful as molecular weight or chromosome
XX markers, to identify genetic disorders, as hybridisation probes or
XX primers, as an antigen, identify and express recombinant protein for
XX analysis, characterisation or therapeutic use, or as markers for tissues
XX in which the corresponding protein is expressed. Diagnostic methods from
XX the present invention can be used to identify subjects having or at risk
XX of developing a disease or disorder associated with aberrant expression
XX or activity of (I). The assays can be utilised to identify a subject
XX having or at risk of developing a disorder associated with Spastin
XX protein or spastin gene expression or activity. The present sequence
XX encodes mouse spastin as given in the present invention.
XX
SQ Sequence 11493 BP; 3599 A; 2281 C; 2387 G; 3226 T; 0 other;

Query Match 92.0%; Score 18.4; DB 22; Length 11493;
Best Local Similarity 95.0%; Pred. No. 6.7;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gtgaatggcaccattgcact 20
|||||
Db 6397 gtgaatggcaccattgcact 6416

RESULT 8
ABA08683
ID ABA08683 standard; cDNA: 901 BP.
XX
AC ABA08683;
XX
DT 11-JAN-2002 (first entry)
XX
DE Human PRO257 homologue-encoding cDNA, SEQ ID NO:459.

XX Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; actvlin;
KW inhibin; chemotaxis; chemokinesis; thrombolytic; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiaesthetic; antiarthritic; haemostatic; antiatherosclerotic;
KW cytosstatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW antifungal; vulnerary; antitumor; ss.
XX
OS Homo sapiens.
XX
PN MO200157188-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US03800.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG15064.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 15055; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 2206 BP; 573 A; 638 C; 481 G; 514 T; 0 other;

Query Match 82.0%; Score 16.4; DB 23; Length 2206;
Best Local Similarity 94.4%; Pred. No. 58;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gtgaatggccacttgcga 18
|||||
Db 2045 GTGAATGGCCACTCTGCA 2028

RESULT 11
AAS82805/c
ID AAS82805 standard; CDNA; 2206 BP.
XX
AC AAS82805;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #18609.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG18618.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 18609; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 2206 BP; 573 A; 638 C; 481 G; 514 T; 0 other;

Query Match 82.0%; Score 16.4; DB 23; Length 2206;
Best Local Similarity 94.4%; Pred. No. 58;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gtgaatggccacttgcga 18
|||||
Db 2045 GTGAATGGCCACTCTGCA 2028

RESULT 12
AAS83970/c
ID AAS83970 standard; CDNA; 2206 BP.
XX
AC AAS83970;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #19774.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.
 XX XX WO200175067-A2.
 XX PN 11-OCT-2001.
 XX PD 30-MAR-2001; 2001WO-US08631.
 XX PF 31-MAR-2000; 2000US-0540217.
 XX PR 23-AUG-2000; 2000US-0649167.
 XX XX (HYSE-) HYSEQ INC.
 XX PA Drmanac RT, Liu C, Tang YT;
 XX PI WPI; 2001-639362/73.
 XX DR P-PSDB; ABG19783.
 XX XX
 XX PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 PS Claim 1; SEQ ID NO 19774; 103bp; English.

XX CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX XX
 XX SQ Sequence 2206 BP; 573 A; 638 C; 481 G; 514 T; 0 other;

Query Match 82.0%; Score 16.4; DB 23; Length 2206;
 Best Local Similarity 94.4%; Pred. No. 58;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gtgaatgacacttgcga 18
 |||||
 DB 2045 GTGAATGCCACTCTGCA 2028

RESULT 13
 AAS85113/c
 ID AAS85113 standard; cDNA; 2206 BP.
 XX

AC AAS85113;
 XX
 XX 13-FEB-2002 (first entry)
 XX

DE DNA encoding novel human diagnostic protein #20917.
 XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW Food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX

OS OS Homo sapiens.
 XX XX WO200175067-A2.
 XX PN 11-OCT-2001.
 XX PD 30-MAR-2001; 2001WO-US08631.
 XX PF 31-MAR-2000; 2000US-0540217.
 XX PR 23-AUG-2000; 2000US-0649167.
 XX XX (HYSE-) HYSEQ INC.
 XX PA Drmanac RT, Liu C, Tang YT;
 XX PI WPI; 2001-639362/73.
 XX DR P-PSDB; ABG20926.
 XX XX
 XX PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 PS Claim 1; SEQ ID NO 20917; 103bp; English.

XX CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX XX
 XX SQ Sequence 2206 BP; 573 A; 638 C; 481 G; 514 T; 0 other;

Query Match 82.0%; Score 16.4; DB 23; Length 2206;
 Best Local Similarity 94.4%; Pred. No. 58;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gtgaatgacacttgcga 18
 |||||
 DB 2045 GTGAATGCCACTCTGCA 2028

RESULT 14
 AAS88179/c
 ID AAS88179 standard; cDNA; 2206 BP.
 XX

AC AAS88179;
 XX
 XX 13-FEB-2002 (first entry)
 XX

DE DNA encoding novel human diagnostic protein #23983.
 XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW Food supplement; medical imaging; diagnostic; genetic disorder; ss.
 OS Homo sapiens.

XX MO200175067-A2.
PN 11-OCT-2001.
XX
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI: 2001-639362/73.
DR P-PSDB; ABG23992.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID No 23983; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2206 BP; 573 A; 638 C; 481 G; 514 T; 0 other;
SQ

Query Match 82.0%; Score 16.4; DB 23; Length 2206;
Best Local Similarity 94.4%; Pred. No. 58;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gtgaatggccacttgcga 18
|||
Db 2045 GTGAATGGCCACTCTGCA 2028

RESULT 15
AAS89592/c
ID AAS89592 standard; cDNA; 2206 BP.
XX
XX AAS89592;
AC
XX 13-FEB-2002 (first entry)
DT
XX
XX DNA encoding novel human diagnostic protein #25396.
DE
XX
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX

PN MO200175067-A2.
XX 11-OCT-2001.
XX
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI: 2001-639362/73.
DR P-PSDB; ABG25405.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID No 25396; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2206 BP; 573 A; 638 C; 481 G; 514 T; 0 other;
SQ

Query Match 82.0%; Score 16.4; DB 23; Length 2206;
Best Local Similarity 94.4%; Pred. No. 58;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gtgaatggccacttgcga 18
|||
Db 2045 GTGAATGGCCACTCTGCA 2028

Search completed: May 22, 2002, 06:46:09
Job time: 8179 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 06:34:29 ; Search time 108.82 Seconds
(without alignments)
45.145 Million cell updates/sec

Title: US-09-693-205-47

Perfect score: 20
Sequence: 1 gtgaatgccactgtcact 20

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.NA:*
1: /cgn2_6/pcdata/1/lna/5B_COMB.seq:*
2: /cgn2_6/pcdata/1/lna/5A_COMB.seq:*
3: /cgn2_6/pcdata/1/lna/6A_COMB.seq:*
4: /cgn2_6/pcdata/1/lna/6B_COMB.seq:*
5: /cgn2_6/pcdata/1/lna/PCrUS_COMB.seq:*
6: /cgn2_6/pcdata/1/lna/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	15.8	79.0	1445	1	US-08-324-533-1	Sequence 1, Appl1
2	15.4	77.0	5105	1	US-08-148-122A-1	Sequence 1, Appl1
3	15.2	76.0	768	2	US-08-799-464A-2	Sequence 2, Appl1
4	15.2	76.0	768	5	PCR-US95-09927-2	Sequence 2, Appl1
5	15.2	76.0	1304	6	5179023-3	Patent No. 5179023
6	15.2	76.0	2351	4	US-08-478-316-4	Sequence 4, Appl1
7	15.2	76.0	2352	4	US-08-478-316-5	Sequence 5, Appl1
8	15.2	76.0	2352	4	US-08-478-316-7	Sequence 7, Appl1
9	15.2	76.0	3293	4	US-08-478-316-40	Sequence 40, Appl1
10	15.2	76.0	3293	4	US-08-478-316-41	Sequence 41, Appl1
11	15.2	76.0	3285	4	US-08-686-968C-5	Sequence 5, Appl1
12	15.2	76.0	3358	2	US-08-799-464A-1	Sequence 1, Appl1
13	15.2	76.0	3358	2	PCR-US95-09927-1	Sequence 1, Appl1
14	14.8	74.0	3342	2	US-08-254-989-1	Sequence 1, Appl1
15	14.8	74.0	5596	4	US-09-078-294-5	Sequence 5, Appl1
16	14.8	74.0	80246	4	US-09-078-294-4	Sequence 4, Appl1
17	14.8	74.0	80595	4	US-09-078-294-3	Sequence 3, Appl1
18	14.4	72.0	2574	4	US-09-142-529-2	Sequence 2, Appl1
19	14.2	71.0	551	2	US-08-647-368A-4	Sequence 4, Appl1
20	14.2	71.0	1194	4	US-08-943-731-188	Sequence 188, App
21	14.2	71.0	2205	4	US-08-840-767-53	Sequence 53, Appl1
22	14.2	71.0	2449	4	US-08-840-767-7	Sequence 7, Appl1
23	14.2	71.0	3660	1	US-08-158-232-42	Sequence 42, Appl1
24	14.2	71.0	3660	1	US-08-611-928-42	Sequence 42, Appl1
25	14.2	71.0	3660	3	US-09-173-891-42	Sequence 42, Appl1
26	14.2	71.0	3796	1	US-08-343-760A-1	Sequence 1, Appl1
27	14.2	71.0	20084	4	US-08-943-731-5	Sequence 5, Appl1

c	28	14	70.0	1716	5	PCR-US96-05320A-541	Sequence 541, App
c	29	13.8	69.0	1236	3	US-09-081-320-37	Sequence 37, Appl
c	30	13.8	69.0	1245	1	US-08-385-186-14	Sequence 14, Appl
c	31	13.8	69.0	1247	2	US-08-647-960-1	Sequence 1, Appl1
c	32	13.8	69.0	1405	2	US-08-555-723B-5	Sequence 5, Appl1
c	33	13.8	69.0	1405	2	US-09-123-465-5	Sequence 5, Appl1
c	34	13.8	69.0	1419	3	US-08-726-807B-1	Sequence 1, Appl1
c	35	13.8	69.0	1419	3	US-09-258-367-1	Sequence 1, Appl1
c	36	13.8	69.0	1419	4	US-09-546-550-1	Sequence 1, Appl1
c	37	13.8	69.0	1419	4	US-09-431-414-1	Sequence 1, Appl1
c	38	13.8	69.0	1419	4	US-09-225-670-1	Sequence 1, Appl1
c	39	13.8	69.0	1425	1	US-08-385-186-12	Sequence 12, Appl
c	40	13.8	69.0	1710	4	US-08-630-915A-189	Sequence 189, App
c	41	13.8	69.0	1978	1	US-08-614-801A-3	Sequence 3, Appl1
c	42	13.8	69.0	2271	4	US-08-891-298-2	Sequence 2, Appl1
c	43	13.8	69.0	2797	2	US-08-555-723B-4	Sequence 4, Appl1
c	44	13.8	69.0	2797	3	US-09-123-465-4	Sequence 4, Appl1
c	45	13.8	69.0	3892	2	US-08-555-723B-3	Sequence 3, Appl1

ALIGNMENTS

RESULT 1
US-08-324-533-1/c
Sequence 1, Application US/08324533
Patent No. 5547870
GENERAL INFORMATION:
APPLICANT: Delta, Asis
APPLICANT: Mehta, Anuradha
APPLICANT: Nataraajan, K.
TITLE OR INVENTION: Oxalate Decarboxylase
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cushman, Darby & Cushman
STREET: 1100 New York Avenue, N.W.
City: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/324,533
CLASSIFICATION: 435
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/985,695
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Kokulis, Paul N.
REGISTRATION NUMBER: 16773
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
TELEX: 6714627 cush
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1445 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-324-533-1

Query Match 79.0%; Score 15.8; DB 1; Length 1445;
Best local Similarity 89.5%; Pred. No. 13;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgaatgaccttgcact 20
|||||
DB 345 TGAATGCCACTTGCCT 327

RESULT 2
US-08-148-122A-1/C
; Sequence 1, Application US/08148122A
; Patent No. 5441885
; GENERAL INFORMATION:
; APPLICANT: Goldberg, Ima
; APPLICANT: Allenza, Paul
; TITLE OF INVENTION: IMPROVED BACTERIAL STRAINS FOR
; TITLE OF INVENTION: BIOREMEDIATION
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AlliedSignal Inc.
; STREET: 50 East Algonquin Road
; CITY: Des Plaines
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60017-5016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: AST Bravo 486/33 DX
; OPERATING SYSTEM: DOS 5.00
; SOFTWARE: Microsoft Word 2.0c
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/148,122A
; FILING DATE: No. 544185ember 5, 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/049,281
; FILING DATE: April 13, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary Jo Boldingh
; REGISTRATION NUMBER: 34,713
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (708) 391-3383
; TELEFAX: (708) 391-3356
; TELEX: N/A
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5105
; TYPE: nucleic acid
; STRANDEDNESS: double strand
; TOPOLOGY: linear
; US-08-148-122A-1

Query Match 77.0%; Score 15.4; DB 1; Length 5105;
Best Local Similarity 94.1%; Pred. No. 27;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 aatggcacttgcact 20
|
DB 4880 AGTGGCCTTGCCT 4864

RESULT 3
US-08-799-464A-2/C
; Sequence 2, Application US/08799464A
; Patent No. 598601
; GENERAL INFORMATION:
; APPLICANT: Murtough, Michael P. et al.
; TITLE OF INVENTION: VR-2332 VIRAL NUCLEOTIDE SEQUENCE AND
; TITLE OF INVENTION: METHODS OF USE
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John M. Collins
; STREET: 2405 Grand Blvd., Suite 400

CITY: Kansas City
STATE: Missouri
COUNTRY: USA
ZIP: 64108
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/799,464A
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/287,941
; FILING DATE: August 5, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Collins, John M.
; REGISTRATION NUMBER: 26,262
; REFERENCE/DOCKET NUMBER: 22907
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (816) 474-9050
; TELEFAX: (816) 474-9057
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 768 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..768
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence=EXPERIMENTAL
; OTHER INFORMATION: /standard_name="VR-2332 ORF 2"
; US-08-799-464A-2

Query Match 76.0%; Score 15.2; DB 2; Length 768;
Best Local Similarity 85.0%; Pred. No. 23;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gtgaatgaccttgcact 20
|
DB 261 GGAATGCTCCTTGCCT 242

RESULT 4
PCT-US95-09927-2/C
; Sequence 2, Application PC/TUS9509927
; GENERAL INFORMATION:
; APPLICANT: Murtough, Michael P.
; TITLE OF INVENTION: VR-2332 VIRAL NUCLEOTIDE SEQUENCE AND
; TITLE OF INVENTION: METHODS OF USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John M. Collins
; STREET: 1101 Walnut, Suite 1400
; CITY: Kansas City
; STATE: Missouri
; COUNTRY: USA
; ZIP: 64106
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/09927
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:

NAME: Collins, John M.
REGISTRATION NUMBER: 26122
REFERENCE/DOCKET NUMBER: 22907
TELECOMMUNICATION INFORMATION:
TELEPHONE: (816) 474-9050
TELEFAX: (816) 474-9057
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 768 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..768
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "VR-2332 ORF 2"
PCT-US95-09927-2

Query Match 76.0%; Score 15.2; DB 5; Length 768;
Best Local Similarity 85.0%; Pred. No. 23;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gtgaatggccacttgact 20
| | | | | | | | | | | | | | | | | | | | | |
Db 261 GGAATGTCACCTTGCACT 242

RESULT 5
5179023-3
PATENT NO. 5179023
APPLICANT: CALHOUN, DAVID H.; COPPOLA, GEORGE
TITLE OF INVENTION: RECOMBINANT a-GALACTOSIDASE A THERAPY
FOR FABRY DISEASE
NUMBER OF SEQUENCES: 4
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/520,312
FILING DATE: 07-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 328,421
FILING DATE: 24-MAR-1989
SEQ ID NO: 3
LENGTH: 1304
5179023-3

Query Match 76.0%; Score 15.2; DB 6; Length 1304;
Best Local Similarity 85.0%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gtgaatggccacttgact 20
| | | | | | | | | | | | | | | | | | | | | |
Db 807 gtgaatggccacttgacct 826

RESULT 6
US-08-478-316-4/C
SEQUENCE 4, Application US/08478316
PATENT NO. 6251397
GENERAL INFORMATION:
APPLICANT: PAUL, PREM S.
APPLICANT: HALBUR, PATRICK G.
APPLICANT: MENG, XIANG-JIN
APPLICANT: MOROZOV, IGOR
TITLE OF INVENTION: POLYNUCLEIC ACIDS ISOLATED FROM A PORCINE
REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS (PRRSV), PROTEINS
ENCODED BY THE POLYNUCLEIC ACIDS, VACCINES BASED ON THE
TITLE OF INVENTION: PROTEINS AND/OR POLYNUCLEIC ACIDS, A METHOD OF PROTECTING A
TITLE OF INVENTION: PIG FROM PRRS AND A METHOD OF DETECTING A PRRSV
NUMBER OF SEQUENCES: 95

CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,316
FILING DATE: 07-JUN-1995

CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/301,435
FILING DATE: 01-SEP-1994

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/131,625
FILING DATE: 05-OCT-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/969,071
FILING DATE: 30-OCT-1992

ATTORNEY/AGENT INFORMATION:
NAME: Lavallee, Jean-Paul M.P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 4625-026-55X CIP

TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220

TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2351 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:

ORGANISM: porcine reproductive and respiratory virus
STRAIN: Iowa
INDIVIDUAL ISOLATE: ISU22
US-08-478-316-4

Query Match 76.0%; Score 15.2; DB 4; Length 2351;
Best Local Similarity 85.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gtgaatggccacttgact 20
| | | | | | | | | | | | | | | | | | | | | |
Db 295 GGAATGTCACCTTGCACT 276

RESULT 7
US-08-478-316-5/C
SEQUENCE 5, Application US/08478316
PATENT NO. 6251397
GENERAL INFORMATION:
APPLICANT: PAUL, PREM S.
APPLICANT: HALBUR, PATRICK G.
APPLICANT: MENG, XIANG-JIN
APPLICANT: MOROZOV, IGOR
TITLE OF INVENTION: POLYNUCLEIC ACIDS ISOLATED FROM A PORCINE
REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS (PRRSV), PROTEI
ENCODED BY THE POLYNUCLEIC ACIDS, VACCINES BASED ON THE
TITLE OF INVENTION: PROTEINS AND/OR POLYNUCLEIC ACIDS, A METHOD OF PROTECTING A
TITLE OF INVENTION: PIG FROM PRRS AND A METHOD OF DETECTING A PRRSV
NUMBER OF SEQUENCES: 95
CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT
 ADDRESSEE: P.C.
 STREET: 1755 S. Jefferson Davis Highway, Suite 400
 CITY: Arlington
 STATE: Virginia
 COUNTRY: U.S.A.
 ZIP: 22202
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/478,316
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/301,435
 FILING DATE: 01-SEP-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/131,625
 FILING DATE: 05-OCT-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/969,071
 FILING DATE: 30-OCT-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Lavalleye, Jean-Paul M.P.
 REGISTRATION NUMBER: 31,451
 REFERENCE/DOCKET NUMBER: 4625-026-55X CIP
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 413-3000
 TELEFAX: (703) 413-2220
 TELEX: 248855 OPAT UR
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2352 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: unknown
 TOPOLOGY: unknown
 MOLECULE TYPE: cDNA
 ORIGINAL SOURCE:
 ORGANISM: Porcine reproductive and respiratory virus
 INDIVIDUAL ISOLATE: VR2332
 US-08-478-316-5

	Query Match	76.0%	Score 15.2:	DB 4,	Length 2352:
	Best Local Similarity	85.0%	Pred. NO. 30:		
	Matches 17, Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0.
Oy	1 gtgaatgcacacttgcact	20			
Db	295 GGGAAATGTCACCTTGCGACT	276			

RESULT 8
 US-08-478-316-7/c
 : Sequence 7, Application US/08478316
 : Patent No. 6251397
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: PAUL, PREM S.
 : APPLICANT: HALBUR, PATRICK G.
 : APPLICANT: MENG, XIANG-JIN
 : APPLICANT: MOROZOV, IGOR
 :
 : TITLE OF INVENTION: POLYNUCLEIC ACIDS ISOLATED FROM A PORCINE
 : TITLE OF INVENTION: REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS (PRRSV), PROTEINS
 : TITLE OF INVENTION: ENCODED BY THE POLYNUCLEIC ACIDS, VACCINES BASED ON THE
 : TITLE OF INVENTION: PROTEINS AND/OR POLYNUCLEIC ACIDS, A METHOD OF PROTECTING A
 : TITLE OF INVENTION: PIG FROM PRRS AND A METHOD OF DETECTING A PRRSV
 : NUMBER OF SEQUENCES: 95
 :
 : CORRESPONDENCE ADDRESS:
 :
 : ADDRESSEE: OBION, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
 : ADDRESSEE: P.C.

STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,316
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/301,435
FILING DATE: 01-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/131,625
FILING DATE: 05-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/969,071
FILING DATE: 30-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: Lavallee, Jean-Paul M.P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 4625-026-55X CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2352 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: porcine reproductive and respiratory virus
STRAIN: Iowa
INDIVIDUAL ISOLATE: ISU79
IS-08-478-316-7

Query Match	76.0%	Score 15.2	DB 4	Length 2352
Best Local Similarity	85.0%	Pred. No. 30		
Matches 17, Conservative	0	Mismatches 3	Indels 0	Gaps 0
QY	1	gtgaatgcacattgcact	20	
db	295	gggaatgtccacttgcact	276	

RESULT 9
 US-08-478-316-40/C
 : Sequence 40, Application US/08478316
 : Patent No. 6251397
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: PAUL, PREM S.
 : APPLICANT: HALBUR, PATRICK G.
 : APPLICANT: MENG, XIANG-JIN
 : APPLICANT: MORZOV, IGOR
 :
 : TITLE OF INVENTION: POLYNUCLEIC ACIDS ISOLATED FROM A PORCINE
 : TITLE OF INVENTION: REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS (PRRSV), PROTEIN
 : TITLE OF INVENTION: ENCODED BY THE POLYNUCLEIC ACIDS, VACCINES BASED ON THE
 : TITLE OF INVENTION: PROTEINS AND/OR POLYNUCLEIC ACIDS, A METHOD OF PROTECTING A
 : NUMBER OF SEQUENCES: 95 PIG FROM PRRS AND A METHOD OF DETECTING A PRRSV
 :
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
 : ADDRESSEE: P.C.
 : STREET: 1755 S. Jefferson Davis Highway, Suite 400

CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,316
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/301,435
FILING DATE: 01-SEP-1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/131,625
FILING DATE: 05-OCT-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/969,071
FILING DATE: 30-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye, Jean-Paul M.P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 4625-026-55X CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 3293 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Porcine reproductive and respiratory virus
STRAIN: Iowa
INDIVIDUAL ISOLATE: ISU79
US-08-478-316-40

Query Match 76.0%; Score 15.2; DB 4; Length 3293;
Best Local Similarity 85.0%; Pred. No. 32;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 gtaatgacacttgact 20
Db 314 GGGAAATGTCACCTTGCACT 295

RESULT 10
US-08-478-316-41/C
Sequence 41, Application US/08478316
Patent No. 6251397
GENERAL INFORMATION:
APPLICANT: PAUL, PREM S.
APPLICANT: HALBUR, PATRICK G.
APPLICANT: MENG, XIANG-JIN
APPLICANT: MOROZOV, IGOR
TITLE OF INVENTION: POLYNUCLEIC ACIDS ISOLATED FROM A PORCINE
TITLE OF INVENTION: REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS (PRRSV), PROTEINS
TITLE OF INVENTION: ENCODED BY THE POLYNUCLEIC ACIDS, VACCINES BASED ON THE
TITLE OF INVENTION: PROTEINS AND/OR POLYNUCLEIC ACIDS, A METHOD OF PROTECTING A
TITLE OF INVENTION: PIG FROM PRRS AND A METHOD OF DETECTING A PRRSV
NUMBER OF SEQUENCES: 95
CORRESPONDENCE ADDRESS:
ADDRESSEE: P.C.
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
STREET: 1755 S. Jefferson Davis Highway, Suite 400
City: Arlington

STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,316
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/301,435
FILING DATE: 01-SEP-1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/131,625
FILING DATE: 05-OCT-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/969,071
FILING DATE: 30-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye, Jean-Paul M.P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 4625-026-55X CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 3293 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Porcine reproductive and respiratory virus
STRAIN: Iowa
INDIVIDUAL ISOLATE: ISU1894
US-08-478-316-41

Query Match 76.0%; Score 15.2; DB 4; Length 3293;
Best Local Similarity 85.0%; Pred. No. 32;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 gtaatgacacttgact 20
Db 314 GGGAAATGTCACCTTGCACT 295

RESULT 11
US-08-686-968C-5/C
Sequence 5, Application US/08686968C
Patent No. 6221361
GENERAL INFORMATION:
APPLICANT: Cochran, Mark D.
APPLICANT: Junker, David E.
TITLE OF INVENTION: Recombinant Swinepox Virus
FILE REFERENCE: 39119-H/JML
CURRENT APPLICATION NUMBER: US/08/686,968C
CURRENT FILING DATE: 1996-07-25
NUMBER OF SEQ ID NOS: 231
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 5
LENGTH: 3295
TYPE: DNA
ORGANISM: Porcine reproductive and respiratory syndrome virus
US-08-686-968C-5

Query Match 76.0%; Score 15.2; DB 4; Length 3295;

Best Local Similarity 85.0%; Pred. No. 32;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ggaatgccacttgacctg 20
1 ||||| ||||| |||||
Db 307 GGAATGTCCTGCGACT 288

RESULT 12
US-08-799-464A-1/c
; Sequence 1, Application US/08799464A
; Patent No. 5998601
; GENERAL INFORMATION:
; APPLICANT: Murtough, Michael P. et al.
; TITLE OF INVENTION: VR-2332 VIRAL NUCLEOTIDE SEQUENCE AND
; TITLE OF INVENTION: METHODS OF USE
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: John M. Collins
; STREET: 2405 Grand Blvd., Suite 400
; CITY: Kansas City
; STATE: Missouri
; COUNTRY: USA
; ZIP: 64108
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/799,464A
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/287,941
; FILING DATE: August 5, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Collins, John M.
; REGISTRATION NUMBER: 26,262
; REFERENCE/DOCKET NUMBER: 22907
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (816) 474-9050
; TELEFAX: (816) 474-9057
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3358 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Arteriviridae (Unclassified)
; STRAIN: VR-2332
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..768
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /standard_name= "VR-2332 ORF2"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 624..1385
; OTHER INFORMATION: /standard_name= "VR-2332 ORF 3"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1169..1701
; OTHER INFORMATION: /standard_name= "VR-2332 ORF 4"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1716..2315
; OTHER INFORMATION: /standard_name= "VR-2332 ORF 5"

FEATURE:
NAME/KEY: misc_feature
LOCATION: 2303..2824
OTHER INFORMATION: /standard_name= "VR-2332 ORF 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2817..3185
OTHER INFORMATION: /standard_name= "VR-2332 ORF 7"
US-08-799-464A-1

Query Match 76.0%; Score 15.2; DB 2; Length 3358;
Best Local Similarity 85.0%; Pred. No. 32;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ggaatgccacttgacctg 20
1 ||||| ||||| |||||
Db 261 GGAATGTCCTGCGACT 242

RESULT 13
PCT-US95-09927-1/c
; Sequence 1, Application PC/TUS9509927
; GENERAL INFORMATION:
; APPLICANT: Murtough, Michael P.
; TITLE OF INVENTION: VR-2332 VIRAL NUCLEOTIDE SEQUENCE AND
; TITLE OF INVENTION: METHODS OF USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: John M. Collins
; STREET: 1101 Walnut, Suite 1400
; CITY: Kansas City
; STATE: Missouri
; COUNTRY: USA
; ZIP: 64106
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/09927
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Collins, John M.
; REGISTRATION NUMBER: 26122
; REFERENCE/DOCKET NUMBER: 22907
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (816) 474-9057
; TELEFAX: (816) 474-9057
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3358 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Arteriviridae (Unclassified)
; STRAIN: VR-2332
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..768
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /standard_name= "VR-2332 ORF2"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 624..1385
; OTHER INFORMATION: /standard_name= "VR-2332 ORF 3"


```
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1169..1701
; OTHER INFORMATION: /standard_name= "VR-2332 ORF 4"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1716..2315
; OTHER INFORMATION: /standard_name= "VR-2332 ORF 5"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2303..2824
; OTHER INFORMATION: /standard_name= "VR-2332 ORF 6"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2817..3185
; OTHER INFORMATION: /standard_name= "VR-2332 ORF 7"
; PCT-US95-09927-1

Query Match          76.0%; Score 15.2; DB 5; Length 3358;
Best Local Similarity 85.0%; Pred. No. 32;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 gtaatggccacttgac 20
Db      261 GGAATGCTCCACTTGGCAGT 242

RESULT 14
US-08-254-989-1
; Sequence 1, Application US/08254989
; Patent No. 5859307
; GENERAL INFORMATION:
; APPLICANT: Mombaerts, Peter
; APPLICANT: Tonegawa, Susumu
; APPLICANT: Johnson, Randall S.
; APPLICANT: Papaioannou, Virginia
; TITLE OF INVENTION: Mutant RAG-1 Deficient Animals Having NO
; TITLE OF INVENTION: Mature B and T Lymphocytes
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.
; ZIP: 30309-4530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/254,989
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/830831
; FILING DATE: 04-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: MIT5783
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 815-6508
; TELEFAX: (404) 815-6555
; INFORMATION FOR SEQ. ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3342 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
```

```
;
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Mouse
; US-08-254-989-1

Query Match          74.0%; Score 14.8; DB 2; Length 3342;
Best Local Similarity 88.9%; Pred. No. 53;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 tgaatggccacttgac 19
Db      2640 TGAATGGCAACTTGTGCC 2657

RESULT 15
US-09-078-294-5/C
; Sequence 5, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: DAVIES COL
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ. ID NOS: 29
; SOFTWARE: Patentin Ver. 2.0
; SEQ. ID NO 5
; LENGTH: 5596
; TYPE: DNA
; ORGANISM: BAC-F2 contig 1
; US-09-078-294-5

Query Match          74.0%; Score 14.8; DB 4; Length 5596;
Best Local Similarity 88.9%; Pred. No. 59;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 gtaatggccacttga 18
Db      4457 GGAATGGCCTTGTGCA 4440
```

Search completed: May 22, 2002, 06:34:32
Job time: 7577 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 05:31:36 ; Search time 3619.39 Seconds
(without alignments)
74.581 Million cell updates/sec

Title: US-09-693-205-47

Perfect score: 20

Sequence: 1 gtgaatggccacttgcact 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estlmu:*
5: em_estlov:*
6: em_estlpl:*
7: em_estlro:*
8: em_hlcl:*
9: gb_estcl:*
10: gb_estl2:*
11: gb_hlcl:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	20	100.0	238	9	AA897178	AA897178 am09e08.s
2	20	100.0	1010	10	BM472954	BM472954 AGENCOURT
3	20	100.0	1083	10	BM476887	BM476887 AGENCOURT
4	18.4	92.0	621	12	A0446973	A0446973 mgxb00021
5	18.4	92.0	705	12	A0255211	A0255211 mgxb00098
C 6	17.4	87.0	299	10	BF944309	BF944309 RC5-NN116
C 7	17.4	87.0	300	10	BF944296	BF944296 RC5-NN116
C 8	17.4	87.0	306	10	BF952284	BF952284 RC5-NN116
C 9	17.4	87.0	311	10	BF944321	BF944321 RC5-NN116
C 10	17.4	87.0	313	10	BF944323	BF944323 RC5-NN116
C 11	17.4	87.0	315	10	BF944314	BF944314 RC5-NN116
C 12	17.4	87.0	316	10	BF952276	BF952276 RC5-NN116
C 13	17.4	87.0	434	10	W60474	W60474 zc98d08.s1
C 14	17.4	87.0	477	10	BB652591	BB652591 UI-M-A00-
C 15	17.4	87.0	610	12	A2295173	A2295173 RPCI-23-1
C 16	17.4	87.0	655	9	BB624324	BB624324 BB624324
C 17	16.8	84.0	199	9	BB072918	BB072918 BB072918

18	16.8	84.0	329	10	BG58614	BG58614 PM4-CT080
19	16.8	84.0	601	10	BG624992	BG624992 pgn1c.pk0
C 20	16.8	84.0	611	9	AA163498	AA163498 zq49n06.r
21	16.8	84.0	618	9	AA410991	AA410991 zv03a10.s
22	16.8	84.0	635	9	BB589330	BB589330 BB589330
23	16.8	84.0	639	10	BI557012	BI557012 603238477
24	16.8	84.0	656	12	A2382874	A2382874 1M0140N17
25	16.8	84.0	712	9	AA916846	AA916846 on10f08.s
26	16.8	84.0	813	10	BG781638	BG781638 SEAU0C001
27	16.8	84.0	825	10	BF261027	BF261027 HVSMEF002
28	16.4	82.0	220	10	BG155981	BG155981 saar79c09
C 29	16.4	82.0	581	10	BM334925	BM334925 MEST130-A
C 30	16.4	82.0	581	12	CT2388	AJ727056 Clona Int
C 31	16.4	82.0	593	10	BF263818	BF263818 HVCEA000
C 32	16.4	82.0	628	10	BI005815	BI005815 MR4-HN005
C 33	16.4	82.0	654	12	BH384785	BH384785 AG-ND-145
C 34	16.4	82.0	968	12	CNS00044A	AL075910 Drosoph11
C 35	16.4	82.0	425	12	AO583449	AO583449 RPCI-11-4
C 36	16	80.0	504	12	A2301231	A2301231 UP.565-11
C 37	16	80.0	627	12	A0708956	HS_2121_A
C 38	16	80.0	730	12	BH373560	BH373560 BOHE45TF
C 39	15.8	79.0	154	12	AQ907033	AQ907033 G5STC0975
40	15.8	79.0	296	10	BG220942	BG220942 RST40741
C 41	15.8	79.0	297	9	AV086027	AV086027 AV086027
C 42	15.8	79.0	312	10	T86720	T86720 yd77e05.g1
C 43	15.8	79.0	358	12	BH228144	BH228144 1006144E0
C 44	15.8	79.0	364	10	BG509627	BG509627 sac57d07
45	15.8	79.0	378	12	AQ923385	AQ923385 RPCI-23-2

ALIGNMENTS

RESULT 1
AA897178 238 bp mRNA linear EST 04-JAN-1999
LOCUS am09e08.s1 Soares_NFL_T-GBC.S1 Homo sapiens cDNA clone
DEFINITION IMAGE:1466342 3', mRNA sequence.
AA897178.1 GI:3033798

ACCESSION AA897178
VERSION AA897178.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 238)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)

JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LNLN ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 847 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 132.
Location/Qualifiers
1. 238

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Soares_NFL_T-GBC.S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBH19W, testis NHT, and B-cell
NCI-GAP_GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of

I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo. "

BASE COUNT 71 a 53 c 45 g 69 t

5', mRNA sequence.
ACCESSION BM476887
VERSION BM476887.1 GI:18525929
KEYWORDS EST
SOURCE human.
ORGANISM Homo sapiens

Query Match 100.0%; Score 20; DB 9; Length 238;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggtgaatggccactttgcact 20
|||||
Db 220 ggtgaatggccactttgcact 201

REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: ATCC

RESULT 2
BM472954 1010 bp mRNA linear EST 05-FEB-2002
LOCUS AGENCOURT_6466106 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5574588
DEFINITION 5', mRNA sequence.
ACCESSION BM472954
VERSION BM472954.1 GI:18521996
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

FEATURES
source 1..1083
Location/Qualifiers

REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: ATCC

CDNA library Preparation: Life Technologies, Inc.
CDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM12275 row: j column: 18
High quality sequence stop: 696.

CDNA library Preparation: Life Technologies, Inc.
CDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM12323 row: h column: 13
High quality sequence stop: 738.

FEATURES
source 1..1010
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="5574588"
/clone_lib="NIH_MGC_88"
/tissue_type="quodanal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: small intestine; Vector: pCMV-SPORT6;
Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
oligo-dT primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."

BASE COUNT 293 a 209 c 191 g 317 t

Query Match 100.0%; Score 20; DB 10; Length 1010;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggtgaatggccactttgcact 20
|||||
Db 275 ggtgaatggccactttgcact 294

RESULT 3
BM476887 1083 bp mRNA linear EST 05-FEB-2002
LOCUS AGENCOURT_6481789 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:555441
DEFINITION

Query Match 100.0%; Score 20; DB 10; Length 1083;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggtgaatggccactttgcact 20
|||||
Db 547 ggtgaatggccactttgcact 566

RESULT 4
A0446973 621 bp DNA linear GSS 08-APR-1999
LOCUS mgxb0002115f CUGI Rice Blast BAC Library Magnaporthe grisea genomic
DEFINITION clone mgxb0002115f, DNA sequence.
ACCESSION A0446973.1 GI:4576110
VERSION GSS.
KEYWORDS Magnaporthe grisea.
SOURCE Magnaporthe grisea
ORGANISM Magnaporthe grisea

REFERENCE Sordariomyces incertae sedis; Magnaporthaceae; Magnaporthe.
AUTHORS Yu,Y., Zhu,H., Boyd,C.A., Gaudette,B., Gayle,A., Kingsbury,R.,
Phillips,K., Sasnowski,M., Wing,R.A. and Dean,R.A.
TITLE A BAC End Sequencing Framework to Sequence the Magnaporthe grisea
Genome

JOURNAL Unpublished (1998)
COMMENT Contact: Dean RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson University, Clemson, SC 29634
Tel: 864 656 5737
Fax: 864 656 4293
Email: rdean@clemson.edu
Seq primer: TAATAGACTACTATAGGG

Class: BAC ends
High quality sequence start: 52
High quality sequence stop: 286.

FEATURES

source

Location/Qualifiers

1. 621

/organism="Magnaporthe grisea"

/strain="70-15"

/db_xref="taxon:148305"

/clone="mgxb0002115f"

/clone_lib="CUGI Rice Blast BAC Library"

/tissue_type="Protoplasts"

/lab_host="E. coli DH10B"

/note="Vector: PBACWICH; Site 1: HindIII; Site 2: HindIII; Rice blast is one of the most devastating fungal diseases of rice world wide. It is a filamentous ascomycete with a haploid genome (n=7) of approximately 40 Mbp. Rice blast is an important model fungal pathogen for studying numerous aspects of the fungal-host interaction. In order to facilitate genome wide analysis, a BAC library containing 9216 clones with an average insert size of 130 kbp was constructed. This library represents greater than 25X genome coverage. High density colony filters are available upon request."

BASE COUNT

154 a 154 c 173 g 135 t 5 others

ORIGIN

Query Match

92.0%; Score 18.4; DB 12; Length 621;

Best Local Similarity 95.0%; Pred. No. 78;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gtgaatgccacttgact 20

| | | | | | | | | | | | | | | | | | | | | |

Db 414 GAGATGCCACTTGGACT 433

RESULT 5

A0255211

LOCUS A0255211 705 bp DNA linear GSS 23-OCT-1998

DEFINITION mgxb0009B20r CUGI Rice Blast BAC Library Magnaporthe grisea genomic

clone mgxb0009B20r. DNA sequence.

ACCESSION

A0255211

VERSION

A0255211.1

KEYWORDS

GSS.

SOURCE

Magnaporthe grisea.

ORGANISM

Magnaporthe grisea.

REFERENCE

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe.

AUTHORS

Yu, Y., Zhu, H., Boyd, C.A., Gaudette, B., Gayle, A., Kingsbury, R., Phillips, K., Saslowski, M., Wing, R.A. and Dean, R.A.

TITLE

A BAC End Sequencing Framework to Sequence the Magnaporthe grisea

JOURNAL

Genome

COMMENT

Unpublished (1998)

JOURNAL

Contact: Dean RA

COMMENT

Clemson University Genomics Institute

JOURNAL

Clemson University

COMMENT

100 Jordan Hall, Clemson University, Clemson, SC 29634

JOURNAL

Tel: 864 656 5737

COMMENT

Fax: 864 656 4293

JOURNAL

Email: rdean@clemson.edu

COMMENT

Seq primer: GGAACACGCTATGACCATG

JOURNAL

Class: BAC ends

COMMENT

High quality sequence stop: 365.

JOURNAL

Location/Qualifiers

FEATURES

1. 705

source

/organism="Magnaporthe grisea"

JOURNAL

/strain="70-15"

COMMENT

/db_xref="taxon:148305"

JOURNAL

/clone="mgxb0009B20r"

COMMENT

/clone_lib="CUGI Rice Blast BAC Library"

JOURNAL

/tissue_type="Protoplasts"

COMMENT

/lab_host="E. coli DH10B"

BASE COUNT

177 a 175 c 199 g 154 t

ORIGIN

Query Match 92.0%; Score 18.4; DB 12; Length 705;

Best Local Similarity 95.0%; Pred. No. 81;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gtgaatgccacttgact 20

| | | | | | | | | | | | | | | | | | | | | |

Db 353 GAGATGCCACTTGGACT 372

RESULT 6

BF944309/c

LOCUS BF944309 299 bp mRNA linear EST 22-JAN-2001

DEFINITION RC5-NN1163-131000-011-B11 NN1163 Homo sapiens cDNA, mRNA sequence.

ACCESSION

BF944309

VERSION

BF944309.1

KEYWORDS

EST.

SOURCE

EST.

ORGANISM

human.

REFERENCE

Homo sapiens

AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

JOURNAL

1 (bases 1 to 299)

COMMENT

Dias Neto, E., Garcia Correa, R., Verjovsky-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed

JOURNAL

sequence tags

COMMENT

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL

20202663

COMMENT

Contact: Simpson A.J.G.

JOURNAL

Laboratory of Cancer Genetics

COMMENT

Ludwig Institute for Cancer Research

JOURNAL

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, São Paulo-SP, Brazil

AUTHORS

Tel: +55-11-27049922

COMMENT

Fax: +55-11-2707001

JOURNAL

Email: asimpson@ludwig.org.br

COMMENT

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL

JOURNAL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC56c2-RC5-NN1163-131000-011-B11&ts=2000-10-13&td=1)

COMMENT

Seq primer: puc 18 forward

JOURNAL

High quality sequence stop: 299.

FEATURES

Location/Qualifiers

source

1. 299

JOURNAL

/organism="Homo sapiens"

COMMENT

/db_xref="taxon:9606"

JOURNAL

/clone_lib="NN1163"

COMMENT

/dev_stage="Adult"

JOURNAL

/note="Organ: nervous system; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORFESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

COMMENT

low stringency conditions."

BASE COUNT 68 a 87 c 93 g 51 t
 ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 299;
 Best Local Similarity 94.7%; Pred. No. 1.9e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 tgaatggccactgtgcaact 20
 |||||||||||
 Db 296 TGAATGGCCACTGTGCACT 278

RESULT 7
 LOCUS BF944296/c 300 bp mRNA linear EST 22-JAN-2001
 DEFINITION RC5-NN1163-131000-011-B04 NN1163 Homo sapiens CDNA, mRNA sequence.
 ACCESSION BF944296
 VERSION BF944296.1 GI:12361571
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 306)
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
 Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

REFERENCE
 AUTHORS

TITLE
 JOURNAL
 MEDLINE
 COMMENT

20202663
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l=RC5&t2=RC5-NN1163-
 131000-011-B04&t3=2000-10-13&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 13
 High quality sequence stop: 300.
 Location/Qualifiers

FEATURES

source

1..300
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NN1163"
 /dev_stage="Adult"
 /note="Organ: nervous.system; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 BASE COUNT 68 a 92 c 90 g 50 t
 ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 300;
 Best Local Similarity 94.7%; Pred. No. 1.9e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 2 tgaatggccactgtgcaact 20
 |||||||||||

Db 291 TGAATGGCCACTGTGCACT 273

RESULT 8
 LOCUS BF952284/c 306 bp mRNA linear EST 22-JAN-2001
 DEFINITION RC5-NN1163-101100-012-D07 NN1163 Homo sapiens CDNA, mRNA sequence.
 ACCESSION BF952284
 VERSION BF952284.1 GI:12369559
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 306)
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
 Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

REFERENCE
 AUTHORS

TITLE
 JOURNAL
 MEDLINE
 COMMENT

20202663
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l=RC5&t2=RC5-NN1163-
 101100-012-D07&t3=2000-11-10&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 7
 High quality sequence stop: 305.
 Location/Qualifiers

FEATURES

source

1..306
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NN1163"
 /dev_stage="Adult"
 /note="Organ: nervous.system; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 BASE COUNT 68 a 92 c 93 g 53 t
 ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 306;
 Best Local Similarity 94.7%; Pred. No. 1.9e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 tgaatggccactgtgcaact 20
 |||||||||||
 Db 287 TGAATGGCCACTGTGCACT 269

RESULT 9
 LOCUS BF944321/c 311 bp mRNA linear EST 22-JAN-2001
 DEFINITION RC5-NN1163-131000-011-G11 NN1163 Homo sapiens CDNA, mRNA sequence.
 ACCESSION BF944321
 VERSION BF944321.1 GI:12361596
 KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 311)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC5<2=RC5-NN1163-131000-011-G1<3=2000-10-13<4=1)
Seq primer: puc 18 forward
High quality sequence start: 17
High quality sequence stop: 311.
Location/Qualifiers
1. 311
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NN1163"
/dev_stage="Adult"
/note="Organ: nervous_normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 68 a 97 c 94 g 52 t
ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 311;
Best Local Similarity 94.7%; Pred. No. 2e+02; 1; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 2 tgaatggcacttgcact 20
|||||
Db 302 TGAATGCCACTGTGCCT 284

RESULT 10
BF944323 313 bp mRNA linear EST 22-JAN-2001
LOCUS RC5-NN1163-131000-011-H03 NN1163 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF944323
ACCESSION BF944323.1 GI:12361598
VERSION
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 313)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

TITLE Simpson,A.J.
JOURNAL Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
MEDLINE Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
COMMENT 20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC5<2=RC5-NN1163-131000-011-H03<3=2000-10-13<4=1)
Seq primer: puc 18 forward
High quality sequence start: 17
High quality sequence stop: 313.
Location/Qualifiers
1. 313
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NN1163"
/dev_stage="Adult"
/note="Organ: nervous_normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 69 a 93 c 96 g 55 t
ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 313;
Best Local Similarity 94.7%; Pred. No. 2e+02; 1; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 2 tgaatggcacttgcact 20
|||||
Db 304 TGAATGCCACTGTGCCT 286

RESULT 11
BF944314 315 bp mRNA linear EST 22-JAN-2001
LOCUS RC5-NN1163-131000-011-D08 NN1163 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF944314
ACCESSION BF944314.1 GI:12361589
VERSION
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 315)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC5&t2=RC5-NN1163-131000-011-D08&t3=2000-10-13&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 2
 High quality sequence stop: 315.
 Location/Qualifiers
 1. 315
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_1lb="NN1163"
 /dev_stage="Adult"
 /note="Organ: nervous, normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 54 a 96 c 92 g 73 t
 ORIGIN
 Query Match 87.0%; Score 17.4; DB 10; Length 315;
 Best Local Similarity 94.7%; Pred. No. 2e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 2 tgaatggccacttgacct 20
 |||||||||
 Db 21 tgaatggccacttgacct 39
 |||||||||
 RESULT 12
 BF952276 316 bp mRNA linear EST 22-JAN-2001
 LOCUS RC5-NN1163-101100-012-D08 NN1163 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BF952276
 ACCESSION BF952276.1 GI:12369551
 VERSION EST.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 316)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Baia, G. S., Simpson, D. H., Brunstein, A., deoliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC5&t2=RC5-NN1163-101100-012-D08&t3=2000-11-10&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 69
 High quality sequence stop: 316.

Location/Qualifiers
 1. 316
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_1lb="NN1163"
 /dev_stage="Adult"
 /note="Organ: nervous, normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 74 a 93 c 97 g 52 t
 ORIGIN
 Query Match 87.0%; Score 17.4; DB 10; Length 316;
 Best Local Similarity 94.7%; Pred. No. 2e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 2 tgaatggccacttgacct 20
 |||||||||
 Db 307 tgaatggccacttgacct 289
 |||||||||
 RESULT 13
 W60474 434 bp mRNA linear EST 07-JUN-1996
 LOCUS zc98d08.s1 Pancreatic Islet Homo sapiens cDNA IMAGE:339183
 DEFINITION 3', mRNA sequence.
 ACCESSION W60474
 VERSION W60474.1 GI:1367235
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 434)
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M. F., Chapell, B., Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacey, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M. B., Tan, F., Thierry-Mieg, J., Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marris, M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)
 97044478
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LNLN; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
 Seq primer: mob.RCGA+ET
 High quality sequence stop: 340.
 Location/Qualifiers
 1. 434
 /organism="Homo sapiens"
 /db_xref="GDB:1264557"
 /db_xref="taxon:9606"
 /clone_1lb="IMAGE:339183"
 /clone_1lb="Pancreatic Islet"
 /tissue_type="Pancreatic Islet"
 /lab_host="SOLR cells (kanamycin resistant)"
 /note="Organ: pancreas; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Reference: Hum Mol Gen 2, 1795 (1993) Takeda et al. Cloned unidirectionally. Primer: 01190 dt.
 -5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor sequence: 5' CTCGACTTTTATTTTATTTT 3' "
 BASE COUNT 98 a 93 c 97 g 146 t

ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 434;
 Best Local Similarity 94.7%; Pred. No. 2.2e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 tgaatgacacttgact 20
 |||||
 Db 301 TGAATGCCACTTGGACT 319

RESULT 14

BE652591/c 477 bp mRNA linear EST 06-SEP-2000
 DEFINITION UI-M-A00-ach-d-01-0-UI.r1 NIH_BMAP_MPG Mus musculus CDNA clone
 UI-M-A00-ach-d-01-0-UI 5', mRNA sequence.

ACCESSION BE652591
 VERSION BE652591.1 GI:9978440

KEYWORDS

EST.
 house mouse.

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 477)
 Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene
 discovery

Genome Res. 6 (9), 791-806 (1996)

JOURNAL

97044477

MEDLINE

97044477

COMMENT

97044477

Contact: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890

Email: MEST@mail.nih.gov

CDNA library Preparation: M.B. Soares Lab Clone distribution:

Researchers may obtain BMAP CDNA clones from RESEARCH GENETICS. It

should be noted that Bento Soares is generating a small number of

additional specialized non-redundant arrays of BMAP CDNA's whose

availability will be considered under appropriate and limited

collaborative arrangements. The following repetitive elements were

found in this CDNA sequence: 1-23, >AT-rich#low-complexity

Seq primer: M13 Reverse.

Location/Qualifiers

1..477

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UI-M-A00-ach-d-01-0-UI"

/clone_1ib="NIH_BMAP_MPG"

/dev_stage="27-32 days"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; The

NIH_BMAP_MPG library is a non-normalized library

constructed from mouse pineal gland. The tag is a string

of 5 nucleotides present between the Not I site and the

oligo-dT track. The library was constructed as described

by Bonaldo, Lennon and Soares. Genome Research 6: 791-806

, 1996. Tissue provided by Ms. Annie Novakovich,

Zivic-Miller Laboratories."

BASE COUNT 149 a 80 c 76 g 172 t

ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 477;
 Best Local Similarity 94.7%; Pred. No. 2.2e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gtgaatgacacttgac 19
 |||||

Db 238 GTGAATGCCACTTGGAC 220

RESULT 15

AZ295173

LOCUS

DEFINITION

AZ295173

PCPI-23-132111.TJ PCPI-23

610 bp DNA linear GSS 27-JUL-2000

PCPI-23-132111.TJ PCPI-23

Mus musculus genomic clone PCPI-23-132111

, DNA sequence.

ACCESSION AZ295173

VERSION AZ295173.1

GI:9536958

GSS.

KEYWORDS

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 610)

Zhao, S., Nierman, W., Feldblum, T., Malek, J., Shatsman, S., Aklnet

, B., Levins, M., McGann, S., Tsagaye, G., Geer, K., Krol, M., de Jong, P.

and Fraser, C.M.

Mouse BAC End Sequences from Library PCPI-23

Unpublished (1999)

Other_GSSs: PCPI-23-132111.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library PCPI-23. For BAC

library availability, please contact Pieter de Jong

(pieterdejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)

or from Resea ch Genetics (info@resgen.com). BAC end page:

http://www.tigr.org/tcb/bac/ends/mouse/bac_end_intro.html

Plate: 132 row: I column: 11

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1..610

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="PCPI-23-132111"

/clone_1ib="PCPI-23"

/sex="Female"

/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1:

EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or

brain genomic DNA was isolated and partially digested

with a combination of EcoRI and EcoRI Methylase. Size

selected DNA was cloned into the pBAC3.6 vector at the

EcoRI sites. The ligation products were transformed into

DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 180 a 135 c 94 g 201 t

ORIGIN

Query Match 87.0%; Score 17.4; DB 12; Length 610;
 Best Local Similarity 94.7%; Pred. No. 2.4e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 tgaatgacacttgact 20
 |||||

Db 447 TGAATGCCACTTGGACT 465

Search completed: May 22, 2002, 05:31:42

Job time: 4102 sec

